AMERICAN JOURNAL OF MATHEMATICS

FOUNDED BY THE JOHNS HOPKINS UNIVERSITY

PUBLISHED UNDER THE JOINT AUSPICES OF

THE JOHNS HOPKINS UNIVERSITY
AND
THE AMERICAN MATHEMATICAL SOCIETY

Volume LXI, Number 4
OCTOBER, 1939

THE JOHNS HOPKINS PRESS
BALTIMORE, MARYLAND
U. S. A.

CONTENTS

PAGE

On multipliers for Lagrange problems. By E. J. McShane, . . 809
On irreducibility of transformations. By G. T. Whyburn, . . 820
Property S_n. By R. L. Wilder, ... 823
On circulation functions. By Vincent C. Poor, ... 833
Further generalizations of the Cauchy integral formula. By Dawson G. Fulton, ... 843
The Plateau problem for minimal surfaces of arbitrary topological structure. By Max Shiffman, ... 853
Units in p-adic algebras. By O. F. G. Schilling, ... 883
The exponential representation of canonical matrices. By John Williamson, ... 897
Metric methods in determinant theory. By Leonard M. Blumenthal, 912
On the irreducibility of certain classes of polynomials. By Benjamin Rosenbaum, ... 923
The Fourier-Stieltjes coefficients of a function of bounded variation. By A. C. Schaeffer, ... 934
A class of monotone functions. By G. Baley Price, ... 941
Asymptotic probability distributions and harmonic curves. By E. K. Haviland, ... 947
On Dirichlet series involving random coefficients. By Philip Hartman, 955
A limit theorem for probability distributions on lattices. By E. R. van Kampen and Aurel Wintner, ... 965
On a family of symmetric Bernoulli convolutions. By Paul Erdös, . 974
Asymptotic distributions and the ergodic theorem. By Philip Hartman and Aurel Wintner, ... 977
On the distribution of the values of real almost periodic functions. By M. Kac, E. R. van Kampen and Aurel Wintner, . . 985
A remark on asymptotic curves. By E. R. van Kampen, . . . 992
The derivative circular congruence-representation of a polygenic function. By Edward Kasner and John De Cicco, . . 995
The canonical lines and the extremals of two invariant integrals. By Ruth B. Rasmusen, ... 1004

The American Journal of Mathematics will appear four times yearly.

The subscription price of the Journal for the current volume is $7.50 (foreign postage 50 cents); single numbers $2.00.

A few complete sets of the Journal remain on sale.

Papers intended for publication in the Journal may be sent to any of the Editors.

Editorial communications may be sent to Dr. A. Cohen at The Johns Hopkins University.

Subscriptions to the Journal and all business communications should be sent to The Johns Hopkins Press, Baltimore, Maryland, U. S. A.

Entered as second-class matter at the Baltimore, Maryland, Postoffice, acceptance for mailing at special rate of postage provided for in Section 1103, Act of October 3, 1917, Authorized on July 3, 1918.

PRINTED IN THE UNITED STATES OF AMERICA
BY J. H. FURST COMPANY, BALTIMORE, MARYLAND

ON MULTIPLIERS FOR LAGRANGE PROBLEMS.*

By E. J. McShane.

The proof of the Lagrange multiplier rule for the solutions of variational problems of Lagrange or Bolza type has been presented in a number of places; for example,[1] Bliss [1, 2], Morse and Myers [4]. The proof of the Weierstrass condition was until recently made only under the rather strong hypothesis of normality on every sub-arc. In 1932 Graves [3] generalized the theorem to apply to all normal problems, and stated a necessary condition involving the $\mathcal{E}$-function which is satisfied by certain anormal problems.

There is always an infinite aggregate of systems $[\lambda_0, \lambda_1(x), \cdots, \lambda_m(x)]$ which serve as Lagrange multipliers to give the analogues of the Euler equations, although this aggregate may reduce merely to the multiples of one such set. The object of the present note is to show that for every minimizing curve this aggregate contains a particularly desirable sub-aggregate, consisting of sets $[\lambda_0, \lambda_1(x), \cdots, \lambda_m(x)]$ with $\lambda_0 \geqq 0$ for which the analogues of the Du Bois-Reymond equations and transversality conditions and the Weierstrass condition all hold; from which it follows that the analogues of the Euler equations, the Clebsch condition and the Weierstrass-Erdmann corner condition must hold. The proof is not widely different from Bliss' proof of the multiplier rule, and makes no use whatever of the concept of normality.

Instead of spending several pages in setting forth the statement of the problem and the preliminary theorems, we shall regard this paper as an addendum to the paper of Bliss[2] [1]. We shall suppose that the reader has that paper at hand, and is familiar with its contents as far as the bottom of page 691. (However, we make no use of §§ 6, 7). The present paper begins at the end of the last complete sentence on page 691; the numbering of our equations begins, therefore, with (47). All page reference and all references to equations with numbers below (47) are to be understood as references to the paper of Bliss. One very slight change, however, is convenient. We shall ask that in the last half of page 691 the subscript $p+1$ be everywhere replaced by an arbitrary integer l.

1. A family of comparison arcs. A set (x, y, Y') is *admissible* if it

* Received April 24, 1939.

[1] Numbers in square brackets refer to the very brief bibliography at the end of this paper.

[2] As a result, our theorem is stated only for Lagrange problems. The extension to Bolza problems offers no difficulty.

belongs to the neighborhood [3] $\mathfrak{R}$ of page 676 and the matrix $\| \phi_{\alpha y'_i}(x, y, Y') \|$ has rank m. Suppose that $[X, Y']$ is such that $(X, y(X))$ is not an end-point or corner of E_{12} and $(X, y(X), Y')$ is admissible. It is possible to adjoin linear functions $\bar{\phi}_\beta(y')$, $\beta = m+1, \cdots, n$ to the functions $\phi_\alpha(x, y, y')$ in such a way that the matrix

$$(47) \qquad \left\| \begin{matrix} \phi_{\alpha y'_i}(X, y(X), Y') \\ \bar{\phi}_{\beta y'_i}(Y') \end{matrix} \right\|$$

is non-singular.

By standard theorems on differential equations, for all values of b in a neighborhood of $(0, \cdots, 0)$ there is a curve $y_i = Y_i(x, b)$ which satisfies the differential equations

$$(48) \qquad \begin{aligned} &\bar{\phi}_\beta(Y'(x, b)) = \bar{\phi}_\beta(Y'), && (\beta = m+1, \cdots, n), \\ &\phi_\alpha(x, Y(x, b), Y'(x, b)) = 0, && (\alpha = 1, \cdots, m), \end{aligned}$$

and which has the initial value

$$(49) \qquad Y_i(X, b) = y_i(X, b).$$

The functions $Y_i(x, b)$ are continuous together with their partial derivatives of first order for x near X and b near $(0, \cdots, 0)$.

Now let e be any small non-negative number. Using the enlarged system of functions ϕ_i of page 678, we write the equations

$$(50) \qquad \begin{aligned} &\phi_\beta(x, y, y') = \phi_\beta(x, y(x, b), y'(x, b)), && (\beta = m+1, \cdots, n). \\ &\phi_\alpha(x, y, y') = 0, && (\alpha = 1, \cdots, m), \end{aligned}$$

Suppose first that E_{12} has no corners. If e is sufficiently small, equation (50) will have a unique set of solutions $\bar{y}_i(x, b, e)$ such that

$$(51) \qquad \bar{y}_i(X - e, b, e) = Y_i(X - e, b), \qquad (i = 1, \cdots, n).$$

This solution will be defined on the interval $(x_1(b), x_2(b))$, and $\bar{y}_i$ and $\bar{y}'_i$ will be continuous together with their partial derivatives of first order as to b and e for (b, e) near $(0, \cdots, 0)$. As already mentioned on page 679, if E_{12} has corners we apply the theorems on differential equations successively to the intervals of x between corners, choosing initial values at the values of x defining corners in such a way that the functions (51) are continuous.

Now we define the curve $y = y(x, b, e)$ by the equations

$$(52) \qquad \begin{aligned} &y(x, b, e) = \bar{y}(x, b, e), \ x_1(b) \leqq x < X - e, \\ &y(x, b, e) = Y(x, b), \ X - e \leqq x < X, \\ &y(x, b, e) = y(x, b), \ X \leqq x \leqq x_2(b). \end{aligned}$$

---

[3] $\mathfrak{R}$ is an open set containing the elements $(x, y(x), y'(x))$ of E_{12}; a set can be admissible without having Y' "near" $y'(x)$.

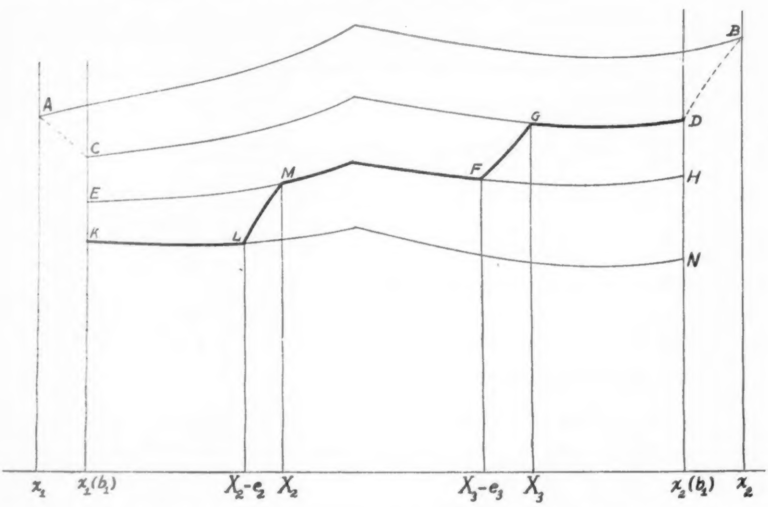
It is evident from the preceding remarks that this is continuous, and that the functions

$$I(b, e) = \int_{x_1(b)}^{x_2(b)} f(x, y(x, b, e), y'(x, b, e))\, dx \tag{53}$$

and

$$\psi_\mu(x_1(b), y(x_1(b), b, e), x_2(b), y(x_2(b), b, e))$$

are continuously differentiable for all b near $(0, \cdots, 0)$ and all small non-negative e. In fact, if we regard the integral $I(b, e)$ in (53) as the sum of three integrals, corresponding respectively to the three subintervals in the definition (52), we see that $I(b, e)$ is continuously differentiable for all (b, e)



AB:	$y = y(x)$	$EFGD$:	$y = y(x, b_1, e_3)$
CD:	$y = y(x, b_1)$	LM:	$y = Y(x, b_1, e_3)$
FG:	$y = Y(x, b_1)$	KLN:	$y = \bar{y}(x, b_1, e_2, e_3)$
EFH:	$y = \bar{y}(x, b_1, e_3)$	$KLMFGD$:	$y = y(x, b_1, e_2, e_3)$

near $(0, 0)$, irrespective of sign. However, if $e < 0$ equations (52) fail to define a single-valued continuous function.

If instead of a single set $[X, Y']$ we have several such sets, say

$$[X_k, Y'_k] \equiv [X'_k, Y'_{1,k}, \cdots, Y'_{n,k}], \quad (k = l + 1, \cdots, s, X_{l+1} < \cdots < X_s),$$

we can iterate the above construction, first using the greatest X_k, then the next greatest, and so on to the least of them. The corresponding parameters e are denoted by e_k, $k = l + 1, \cdots, s$. The resulting type of curve is shown by the heavy curve in the figure, which is drawn for the case $l = 1$, $s = 3$.

The partial derivatives of I with respect to the b_i, at $b = e = 0$, have already been computed, and are expressed in equation (45). If we introduce the notation

$$\eta_{i,k}(x) = \partial \bar{y}_i(x, b, e)/\partial e_k, \quad (b = e = 0;\ k = l+1, \cdots, s, x_1 \leqq x \leqq x_2),$$

we readily find from equations (50) and (11) that the corresponding functions ζ_k vanish identically. Computing the partial derivative of I with respect to e_k for $b = e = 0$, we find

$$(54) \quad \partial I(b, e)/\partial e_k = - f(X_k, y(X_k), y'(X_k)) + \int_{x_1}^{X_k} (f_{y^i}\eta_{i,k} + f_{y'^i}\eta'_{i,k})\,dx + f(X_k, y(X_k), Y'_k).$$

By (13) and (18) this is transformed into

$$(55) \quad \lambda_0 \partial I(b, e)/\partial e_k = -\lambda_0 f(X_k, y(X_k), y'(X_k)) - c_i\eta_{i,k}(x_1) + \eta_{i,k}(X_k) F_{y'_i}(x_k, y(X_k), y'(X_k), \lambda) + F(X_k, y(X_k), Y'_k, \lambda),$$

not summed on k; the left member is understood to be evaluated at $b = e = 0$. If we differentiate both members of (51) with respect to e and set $b = e = 0$, we obtain (interpreting e as any one e_k)

$$-\bar{y}'_i(X_k, 0, 0) + \eta_{ik}(X_k) = - Y'_i(X_k, 0).$$

Since $\bar{y}_i(x, 0, 0) \equiv y_i(x)$ and $Y'_i(X_k, 0) = Y'_{ik}$, this becomes

$$\eta_{ik}(X_k) = -(Y'_{ik} - y'_i(X_k)).$$

Substituting this in equation (55) yields

$$(56) \quad \lambda_0 \partial I(b, e)/\partial e_k = - c_i\eta_{i,k}(x_1) + \mathcal{E}(X_k, y(X_k), y'(X_k), Y'_k, \lambda),$$

where as usual we have written

$$\mathcal{E}(x, y, y', Y', \lambda) = F(x, y, Y', \lambda) - F(x, y, y', \lambda) - (Y'_i - y'_i) F_{y'_i}(x, y, y', \lambda).$$

Observe that the right member of (56) is a continuous function of X_k.

2. A convex set defined by the variations. Next we change notation slightly; we replace the symbol e_k by b_k, $k = l+1, \cdots, s$. We define

$$(57) \quad \begin{aligned} \rho_0(b) &= I(b) - I(0), \\ \rho_\mu(b) &= \psi_\mu(x_1(b), y(x_1(b), b), x_2(b), y(x_2(b), b)), \\ & \qquad (\mu = 1, \cdots, p). \end{aligned}$$

As remarked just after equation (53), these can be regarded as defined, continuous and continuously differentiable for all b near zero, although we must recall that in order that the parameter b shall define a curve of the family $y(x, b)$, all the parameters $b_{l+1}, \cdots, b_s$ must be non-negative.

Let K be the set of points $(u_0, \cdots, u_p)$ in $p+1$-dimensional space defined by the equations

$$(58) \quad u_j = \rho_{ja} b_a, \quad b_a \geqq 0, \quad (\alpha = 1, \cdots, s),$$

wherein

$$\rho_{ja} = \left.\frac{\partial \rho_j(b)}{\partial b_a}\right|_{b=0}.$$

It is easily seen that K is a convex point set; in fact, it is the linear image of the convex set $b_1 \geqq 0, \cdots, b_s \geqq 0$, by (58). We shall need the following lemma.

No point of the negative u_0-axis is interior to the set K.

Suppose there were such a point $\bar{u} : \bar{u}_0 < 0,\ \bar{u}_1 = \cdots = \bar{u}_p = 0$. Since $\bar{u}$ is in K it can be represented by equations

$$\bar{u}_i = \rho_{i\alpha}\bar{b}_\alpha, \qquad \bar{b}_\alpha \geqq 0, \qquad (i = 0, \cdots, p,\ \alpha = 1, \cdots, s). \tag{59}$$

We first show that the numbers $\bar{b}_\alpha$ in (59) may be supposed actually positive. The point $\bar{u}$ is interior to K; so if ϑ is a sufficiently small positive number, the point

$$u'_i = \rho_{i\alpha}b'_\alpha, \qquad b'_\alpha = \bar{b}_\alpha + \vartheta > 0$$

is also interior to K. So is $u'' = 2\bar{u} - u'$; hence

$$u_i'' = \rho_{i\alpha}b_\alpha'', \qquad b_\alpha'' \geqq 0.$$

Thus

$$\bar{u}_i = \tfrac{1}{2}(u_i' + u_i'') = \rho_{i\alpha}[\tfrac{1}{2}(b_\alpha' + b_\alpha'')],$$

and here the coefficients of the $\rho_{i\alpha}$ are positive. If the rows of the matrix $\rho_{i\alpha}$ were linearly dependent, the coördinates u_i of the points of K would satisfy a linear relationship, and K could have no interior points. Hence, as we are supposing $\bar{u}$ interior to K, the rows are linearly independent. Consequently we can adjoin $s - (p+1)$ linear functions $\rho_h(b) = \rho_{h\alpha}b_\alpha,\ h = p+1, \cdots, s-1$, in such a way that the square matrix $\| \rho_{i\alpha} \|$ is non-singular. Consider now the equations

$$u_i - \rho_i(b) = 0, \qquad (i = 0, \cdots, s-1). \tag{60}$$

These are satisfied if $u = b = 0$, and the jacobian with respect to the b_i is non-vanishing. Hence they have solutions $b_i = b_i(u),\ i = 1, \cdots, s$, which are defined and continuous and possess continuous first partial derivatives near $u = 0$, and are such that $b_i(0) = 0$.

The numbers $\bar{u}_i,\ i = 0, \cdots, p$ are defined by equation (59). We augment this set, defining

$$\bar{u}_i = \rho_{i\alpha}\bar{b}_\alpha, \qquad (i = 0, \cdots, s-1). \tag{61}$$

The equation

$$t\bar{u}_i = \rho_i(b(t\bar{u})) \tag{62}$$

is an identity by definition of the function $b(u)$. If we differentiate with respect to t and set $t = 0$, we obtain

$$\bar{u}_i = \rho_{i\alpha}\frac{db_\alpha(t\bar{u})}{dt}.$$

The matrix $\| \rho_{i\alpha} \|$ is non-singular, so this and (61) imply

$$\frac{db_a(t\bar{u})}{dt} = \bar{b}_a > 0.$$

Hence for all small positive t the functions $b_a(t\bar{u})$ are all positive-valued. Also, if t is small the curve $y = y(x, b)$ is in an arbitrarily small neighborhood of E_{12}. Equation (62), with (57) and the fact that $(\bar{u}_0, \cdot \cdot \cdot, \bar{u}_p)$ is on the negative u_0-axis, yields

$$\begin{aligned} &I(b(t\bar{u})) < I(0), \\ &\psi_\mu(X_1(b(t(\bar{u})), y(x_1(b(t\bar{u})), b(t\bar{u})), x_2(b(t\bar{u}))), \\ &\qquad\qquad\qquad\qquad y(x_2(b(t\bar{u})), b(t\bar{u})) = 0. \end{aligned}$$

The curves $y = y(x, b)$ were constructed so as to satisfy the differential equations $\phi_a = 0$. Now we see that if $b = b(t\bar{u})$, t small and positive, they also satisfy the end conditions $\psi_\mu = 0$ and give I a smaller value than $I(0)$. This contradicts the hypothesis that E_{12} is a minimizing curve, and our lemma is established.

The set K depends for its definition on the sets $[X_k, Y'_k]$ and $[\xi_{i1}, \xi_{i2}, \eta_{ij}(x)]$ used in defining the functions $y(x, b)$. Let K^* be the closure of the sum of all sets K, for all possible choices of sets $[X_k, Y'_k]$ and $[\xi_i, \eta_i]$. That is, u is in K^* if every neighborhood of u contains points belonging to some set K, as defined by (58). We first prove

The set K^* is convex.

Let $v, \tilde{v}$ be two points in K^*. For every positive number ϵ there are points $u, \tilde{u}$ having distances less than ϵ from $v, \tilde{v}$ respectively and belonging to sets $K, \tilde{K}$. Let $[X_k, Y'_k]$ and $[\xi_i, \eta_i]$ be the sets defining K, and $[\tilde{X}_h, Y'_h]$ and $[\tilde{\xi}_j, \tilde{\eta}_j]$ the sets defining $\tilde{K}$. Then u is defined by (58), and $\tilde{u}$ analogously. We may suppose that the points $\tilde{X}_h$ and X_k are all distinct, since by the remark after equation (56) the X_k can be moved slightly so as to produce an arbitrarily small change in the $\rho_{t\alpha}$. If we now use all the sets $[X_k, Y'_k]$, $[\tilde{X}_h, \tilde{Y}'_h]$, $[\xi_i, \eta_i]$, $[\tilde{\xi}_j, \tilde{\eta}_j]$ to define a new set K_1 as in (58), the set K_1 contains both K and $\tilde{K}$, hence contains u and $\tilde{u}$. Being convex, it contains the line segment joining u and $\tilde{u}$. Every point of the line segment joining v and $\tilde{v}$ has distance less than ϵ from some point of the segment joining u and $\tilde{u}$. Since ϵ is arbitrary, every point of the line segment joining v and $\tilde{v}$ is in K^*, and K^* is convex.

No point of the negative u_0-axis is interior to K^*.

Suppose there is such a point $\bar{u}$. It is then possible to find $p + 2$ points $v^{(1)}, \cdot \cdot \cdot, v^{(p+2)}$ in K^* such that $\bar{u}$ is interior to the simplex with vertices

$v^{(1)}, \cdots, v^{(p+2)}$. Arbitrarily near each $v^{(i)}$ there is a point $u^{(i)}$ belonging to some set $K^{(i)}$ defined by sets $[X_k^{(i)}, Y_k^{(i)\prime}]$ and $[\xi_j^{(i)}, \eta_j^{(i)}]$. As before, we may suppose that all the numbers $X_k^{(i)}$ are distinct, and combine all the sets into a single set. This aggregate of sets then defines a convex point set K_1 containing each of the $K^{(i)}$. In particular, it contains the entire simplex with vertices $u^{(i)}$; and if the $u^{(i)}$ are near enough to the $v^{(i)}$, the point $\bar{u}$ is interior to this simplex. But now $\bar{u}$ is interior to K_1, contradicting a preceding lemma.

As K^* does not consist of the entire u-space, it has frontier points. Let $\bar{u}$ be a frontier point of K^*. Since K^* is convex, there is a hyperplane of support[4] of K^* passing through $\bar{u}$. That is, there are numbers $\lambda_0, d_1, \cdots, d_p, q$ such that

$$\lambda_0 u_0 + d_\mu u_\mu + q \geqq 0 \text{ for all } u \text{ in } K^*,\qquad (63)$$
$$\lambda_0 \bar{u}_0 + d_\mu \bar{u}_\mu + q = 0.$$

It is easy to see that q must vanish. If $\bar{u} = (0, \cdots, 0)$ this is obvious from the second of conditions (63). Otherwise, we observe that the points $u = (0, \cdots, 0)$ and $u = 2\bar{u}$ both are in K^*. Substituting these in the first of conditions (63) and using the second yields $q = 0$.

We now wish to show that it is possible to choose λ_0 and the d_i in such a way that λ_0 is non-negative. If $\bar{u} = (-1, 0, \cdots, 0)$ is a frontier point of K^*, we choose λ_0, d_μ so as to define a hyperplane of support at $\bar{u}$. By the second of conditions (63) we obtain $\lambda_0 = 0$. If $(-1, 0, \cdots, 0)$ is not a frontier point of K^*, it is exterior to K^*. There is then a hyperplane separating[5] $(-1, 0, \cdots, 0)$ from K^*. We choose the notation so that

$$\lambda_0 u_0 + d_\mu u_\mu + q > 0 \text{ for all } u \text{ in } K^*,\qquad (64)$$
$$\lambda_0(-1) + d_\mu 0 + q < 0.$$

The first of these, with $u = (0, \cdots, 0)$, implies $q > 0$; the second then implies $\lambda_0 > 0$. For this λ_0 and d_μ the first of conditions (63) holds with q replaced by zero; if it failed for some $\bar{u}$, then for a sufficiently large number N the numbers $u_i = N\bar{u}_i$ would violate (64). Hence $\lambda_0 u_0 + d_\mu u_\mu = 0$ is a hyperplane supporting K^* at the origin.

Summing up our lemmas, we have the following statement.

There are numbers $\lambda_0 \geqq 0$, $d_1, \cdots, d_p$, not all zero, such that for every finite collection of sets $[X_k, Y'_k]$ with $(X_k, y(X_k), Y'_k)$ admissible[6] and every finite collection of sets of admissible variations $[\xi_i, \eta_i]$ the inequalities

---

[4] C. Carathéodory, "Über den Variabilitätsbereich der Fourier'schen Konstanten von positiven harmonischen Funktionen," *Rend. Cir. Math. Palermo*, vol. 32 (1911), pp. 197, 198.

[5] C. Carathéodory, *loc. cit.*

[6] We still suppose that $(X_k, y(X_k))$ is not an end or corner of E_{12}.

$$(65) \qquad [\lambda_0 \rho_{0\beta} + d_\mu \rho_{\mu\beta}] b_\beta \geqq 0$$

hold whenever the numbers b_β are all non-negative.

3. The multiplier rule and the Weierstrass condition. Now let us suppose that there are no sets $[X_k, Y'_k]$ and just one set of variations $[\xi, \eta]$. The matrix $\| \rho_{i\alpha} \|$ has then a single column $\alpha = 1$. Inequality (65) takes the form

$$(66) \qquad \lambda_0 \rho_{0,1} + d_\mu \rho_{\mu,1} \geqq 0.$$

However, $[-\xi, -\eta]$ is also an admissible set of variations, and for these the matrix $\| \rho_{i\alpha} \|$ consists of a single column $(-\rho_{0,1}, \cdot \cdot \cdot, -\rho_{p,1})$; hence by (65) we have

$$\lambda_0(-\rho_{0,1}) + d_\mu(-\rho_{\mu,1}) \geqq 0.$$

This, with (66), implies

$$\lambda_0 \rho_{0,1} + d_\mu \rho_{\mu,1} = 0.$$

In this equation we substitute for the derivatives $\rho_{i,1}$ their values as computed from (57); we obtain

$$\lambda_0 I_1 + d_\mu \Psi_\mu(\xi, \eta) = 0.$$

With the help of equations (44) and (45) this becomes

$$(67) \qquad \begin{aligned} -\int_{x_1}^{x_2} \lambda_r \zeta_r dx &+ [-\lambda_0 f(x_1) + d_\mu(\psi_{\mu x_1} + \psi_{\mu y'_i} y'_{i1})] \xi_1 \\ &+ [\lambda_0 f(x_2) + d_\mu(\psi_{\mu x_2} + \psi_{\mu y_{i2}} y'_{i2})] \xi_2 \\ &+ [-c_i + d_\mu \psi_{\mu y_{i1}}] \eta_i(x_1) \\ &+ [F_{y'_i}(x_2) + d_\mu \psi_{\mu y_{i2}}] \eta_i(x_2) = 0. \end{aligned}$$

The constants λ_0, c_i have so far been arbitrary, save that the λ_0 has recently been chosen in the lemma at the end of § 2. We now choose the c_i so that the coefficients of the $\eta_i(x_1)$ vanish:

$$(68) \qquad c_i = d_\mu \psi_{y_{i1}}.$$

The ξ_1 and ξ_2 are arbitrary, so if (67) is to hold their coefficients must vanish. Likewise, for each set ζ_r the solution $\eta_i(x)$ of equation (11) can be chosen so as to have arbitrary values at x_2; hence the coefficient of $\eta_i(x_2)$ must vanish. Finally, the ζ_r are arbitrary continuous functions, so $\lambda_r(x)$ must vanish identically, $r = m + 1, \cdot \cdot \cdot, n$.

The conditions on the various coefficients imply that the rank of the matrix

$$(69) \qquad \left\| \begin{matrix} -\lambda_0 f(x_1) & -c_i & \lambda_0 f(x_2) & F_{y'_i}(x_2) \\ \psi_{\mu x_1} + \psi_{\mu y_{i1}} y'_{i1} & \psi_{\mu y_{i1}} & \mu x_2 + \mu y_{i2} y'_{i2} & \psi_{\mu y_{i2}} \end{matrix} \right\|$$

is at most p, for if the rows be multiplied by $1, d_1, \cdot \cdot \cdot, d_p$ respectively and

added, the sum is a row of zeros. By (15) we have $c_i = F_{y_i'}(x_1)$. We substitute this in the matrix. Adding a multiple of one column of a matrix to another column leaves the rank unchanged, so the matrix

$$(70) \qquad \begin{Vmatrix} -F(x_1) + y'_{i1}F_{y'_i}(x_1) & -F_{y'_i}(x_1) & F(x_2) - y'_{i2}F_{y'_i}(x_2) & F_{y'_i}(x_2) \\ \psi_{\mu x_1} & \psi_{\mu y_{i1}} & \psi_{\mu x_2} & \psi_{\mu y_{i2}} \end{Vmatrix}$$

has rank less than $p + 1$.

Next, let us use a single set $[X_1, Y'_i]$ and no sets $[\xi, \eta]$. Inequality (65) now takes the form (by (56) and (57))

$$\begin{aligned} &\mathcal{E}(X_1, y(X_1), y'(X_1), Y'_1, \lambda) + c_i\eta_i(x_1) \\ &\qquad + d_\mu\Psi_\mu(0, 0, \eta(x_1), 0) \geqq 0. \end{aligned}$$

If we use (68) and (44), this yields

$$\mathcal{E}(X_1, y(X_1), y(X_1), Y', \lambda) \geqq 0.$$

Collecting the various statements established above and adding two minor ones yet to be proved, we have the following theorem.

For every minimizing arc for the problem of Lagrange with variable end points there exists a non-negative constant λ_0 and a set of functions $\lambda_1(x), \cdots, \lambda_m(x)$ such that for the function

$$F(x, y, y', \lambda) \equiv \lambda_0 f + \lambda_1\phi_1 + \cdots + \lambda_m\phi_m$$

the following statements hold.

(i) (Du Bois-Reymond relation).

There are constants $c_1, \cdots, c_n$ such that the equations

$$F_{y_i'}(x, y(x), y'(x), \lambda) = \int_{x_1}^{x} F_{y_i} dx + c_i$$

hold on the entire interval $[x_1, x_2]$.

(ii) (Transversality).

The rank of the matrix (70) is less than $p + 1$.

(iii) (Weierstrass Condition).

For all x in the interval $[x_1, x_2]$ and all Y' such that $(x, y(x), Y')$ is admissible, the inequality

$$\mathcal{E}(x, y(x), y'(x), Y', \lambda) \geqq 0$$

is satisfied.

(iv) (Clebsch Condition).

For all x in the interval $[x_1, x_2]$ and all sets of numbers $\pi_1, \cdots, \pi_n$ satisfying the equations

$$\pi_i \phi_{\alpha y'_i} = 0, \qquad (\alpha = 1, \cdots, m)$$

the inequality

$$\pi_i F_{y'_i y'_j} \pi_j \geqq 0$$

is satisfied.

Moreover, the constant λ_0 and the functions $\lambda_\alpha(x)$ are not all identically zero on $[x_1, x_2]$, and are continuous except possibly at values of x defining corners of E_{12}.

The continuity properties of the $\lambda_\alpha(x)$ have already been established. If λ_0 and the $\lambda_\alpha(x)$ all were identically zero, the first row of matrix (69) would be identically zero. The numbers λ_0, d_μ are not all zero; since $\lambda_0 = 0$, not all the d are zero. But we have seen that if the rows of (69) are multiplied by $1, d_1, \cdots, d_p$ respectively and added, the sum is a row of zeros. If the first row vanished, we would have a linear dependency (with coefficients d_μ) between the remaining rows, and the rows of matrix (42) would be linearly dependent, contrary to hypothesis.

We have established the Weierstrass condition only under the assumption that $(x, y(x))$ is neither a corner nor an end of E_{12}. It extends to such points readily, by simple continuity considerations; at a corner, we can understand $y'(x)$ to mean either the right or the left derivative.

The condition of Clebsch is a consequence of the condition of Weierstrass, as shown on page 718.

The conclusion that $\lambda_0, \lambda_\alpha(x)$ are not all identically zero can be sharpened to the form that for no x in the interval (x_1, x_2) is $(\lambda_0, \lambda_1(x), \cdots, \lambda_m(x))$ equal to $(0, 0, \cdots, 0)$; cf. [II], p. 27. Furthermore [II, p. 31] we can add one more equation analogous to the Du Bois-Reymond equations:

$$F - y'_i F_{y'_i} = c_0 + \int_{x_1}^{x} F_x dx. \tag{71}$$

4. Remarks on the determination of the multipliers. If the minimizing curve E_{12} happens to be normal, we can take $\lambda_0 = 1$, and then there exists a unique set of functions $\lambda_\alpha(x)$ for which the Du Bois-Reymond equations and the transversality conditions hold. By the theorem just stated, for these same multipliers the Weierstrass and Clebsch conditions must hold.

If E_{12} is not normal, there may be many determinations of λ_0 and the $\lambda_\alpha(x)$ for which the Du Bois-Reymond equations and the transversality conditions hold. For at least one, and possibly for several, of these determinations, the Weierstrass and Clebsch conditions will also hold. However, these latter conditions may impose additional restrictions on the choice of the λ_0 and $\lambda_\alpha(x)$. For example, let $x_1 = 0, x_2 = 1, f(x, y, y') = (y'_i y'_i)^2$. We suppose

the end-values fixed at zero, and take a single side-equation $\phi(x, y, y') = 0$. If we choose $\phi = y'_1 + (y'_1)^3$, then for any λ_0 and any constant $\lambda(x)$ the Du Bois-Reymond relations hold along the arc $E_{12}: y = 0$. But unless $\lambda(x)$ is identically zero the Weierstrass condition is not satisfied. Again, if we choose $\phi = y'_1 + y'_i y'_i$, for any λ_0 and any constant $\lambda(x)$ the Du Bois-Reymond relations hold. But for the Weierstrass condition to hold we must have $\lambda(x) \geqq 0$.

It is possible to establish the Jacobi condition by considerations similar to the preceding, but not without assumptions of normality; the method seems at present to be useful only for normal problems and certain problems with order of anormality 1. The inclusion of the proof of the Jacobi condition for such cases would about double the length of this paper, and the theorem does not seem to be interesting enough to justify the use of the extra pages.

If the necessary condition of Jacobi could be established without assumptions of normality, we would be in possession of a complete set of necessary conditions and sufficient conditions for a minimum without hypotheses of normality. It would be rash to conclude that the concept of normality would be rendered useless. For within the class of anormal problems there are problems of definitely unruly behavior, such as those in which there is but a single curve satisfying the differential equation and end conditions. The distinction between such problems and those of more placid aspect is intrinsic, and not erased by any amount of analytic ingenuity. Nevertheless, it seems desirable to have proofs constructed so as to apply simultaneously both to normal and to anormal problems, if for no other reason than to avoid having to verify the presence of normality in individual cases where such verification may be difficult.

UNIVERSITY, VA.

REFERENCES.

[1] G. A. Bliss, "The problem of Lagrange in the calculus of variations," *American Journal of Mathematics*, vol. 52 (1930), pp. 673-744.

[2] G. A. Bliss, Mimeographed notes of lectures on the problem of Bolza, delivered at the University of Chicago, 1935.

[3] L. M. Graves, "On the Weierstrass condition for the problem of Bolza in the calculus of variations," *Annals of Mathematics*, vol. 33 (1932), pp. 747-752.

[4] Marston Morse and S. B. Myers, "The problems of Lagrange and Mayer with variable end points," *Proceedings of the American Academy of Arts and Sciences*, vol. 66 (1931), pp. 235-253.

ON IRREDUCIBILITY OF TRANSFORMATIONS.*

By G. T. Whyburn.

If A is a compact metric continuum, a continuous transformation $T(A) = B$ has been called [1] *irreducible* provided no proper subcontinuum of A maps onto all of B under T. Similarly, merely assuming A compact, T is said [2] to be *strongly irreducible* provided no closed proper subset of A maps onto all of B under T.

In this paper certain auxiliary non-negative real valued functions will be defined for such a transformation T on a compact set A and application will be made of the continuity properties of these functions to yield certain conclusions about T, in particular about irreducibility properties of T.

Let A be compact and metric and let $T(A) = B$ be continuous. We define

$$f(x) = \delta(T^{-1}T(x)), \qquad x \in A;$$
$$g(x) = \delta(C_x), \text{ where } C_x \text{ is the component of } T^{-1}T(x) \text{ containing } x;$$
$$h(x) = \text{l. u. b. } [\delta(C)], \text{ where } C \text{ is a component of } T^{-1}T(x).$$

It is readily seen that the functions f, g and h are upper semi-continuous. Since they are non-negative, they therefore are continuous on the sets D_f, D_g, D_h respectively where they vanish. Now if E_f, E_g, E_h denote the sets on which f, g, and h respectively are continuous, we have that D_f, D_g, D_h are relatively closed subsets of E_f, E_g, E_h respectively. Thus since E_f, E_g, E_h are G_δ-sets, so also are D_f, D_g, and D_h.

Thus we have

Theorem 1. *If A is compact and $T(A) = B$ is continuous, the sets*
D_f of all points of A on which T is $(1-1)$,
D_g of all points x such that x is a component of $T^{-1}T(x)$,
D_g of all points where T is light (i. e., *all points x such that $T^{-1}T(x)$ is 0-dimensional*) *are G_δ-sets.*

In a similar way, if we define $n(x)$ to be the number of components of $T^{-1}T(x)$ and assume T interior, it follows that $n(x)$ is lower semi-continuous on A. Thus for any integer n, the set E_n of all points x of A with $n(x) \leq n$

* Received July 6, 1939.

[1] See my paper in the *American Journal of Mathematics*, vol. 58 (1936), p. 305.

[2] See J. Rozanska, *Fundamenta Mathematicae*, vol. 28 (1937), p. 226, and R. G. Simond, *Duke Mathematical Journal*, vol. 4 (1938), p. 578.

is closed. In particular, since $n(x) \geq 1$ everywhere on A, the set E_n of all points x of A with $n(x) = 1$, i. e., the subset of A on which T is monotone, is closed. Of course, this is a conclusion easily arrived at directly without the use of the function $n(x)$.

THEOREM 2. *In order that a continuous transformation $T(A) = B$ on a compact set A be strongly irreducible it is necessary and sufficient that the set D_f of all points $x \epsilon A$ with $x = T^{-1}T(x)$ be dense in A.*

Proof. The condition clearly is sufficient, since obviously any subset of A which maps onto all of B must contain D_f. To establish the necessity of the condition, let O be any open subset of A. Then since T is strongly irreducible there must exist at least one point x such that $T^{-1}T(x) \subset O$, because $T(A - O) \neq B$. Thus for any $\epsilon > 0$, the set E_ϵ of all points $x \epsilon A$ with $f(x) < \epsilon$ is dense in A; and since $f(x)$ is upper semi-continuous, it follows that the set D_f of all points $x \epsilon A$ where $f(x) = 0$ is also dense in A.

COROLLARY. *If $T(A) = B$ is strongly irreducible and interior, where A is compact, then T is topological.*

THEOREM 3. *Let A be a compact semi-locally-connected continuum.*[3] *In order that a continuous transformation $T(A) = B$ be irreducible it is necessary and sufficient that the set D_f of all points $x \epsilon A$ with $x = T^{-1}T(x)$ be dense on the set N of all non-cut points of A.*

Proof. The condition is sufficient. For suppose on the contrary that some proper subcontinuum K of A maps onto all of B under T. Then clearly we have $D_f \subset K$. But this gives $N \subset K$, since $\bar{D}_f \supset N$. Thus any $x \epsilon A - K$ is a cut point; and this is impossible since any component of $A - x$ necessarily contains at least one non-cut point of A. To prove the necessity of the condition we again make use of the function $f(x)$. For any open set O in A intersecting $\bar{N}$, there exists at least one x such that $T^{-1}T(x) \subset O$, because about a point of $N \cdot O$ we can get[3] an open set O_1 such that $O_1 \subset O$ and $A - O_1$ is a continuum and $T(A - O_1) \neq B$. Thus for any $\epsilon > 0$ the set E_ϵ of all points $x \epsilon A$ with $f(x) < \epsilon$ is dense in $\bar{N}$; and since $f(x)$ is upper semi-continuous, the set D_f of all points $x \epsilon A$ with $f(x) = 0$ must be dense in $\bar{N}$.

---

[3] A connected metric set M is semi-locally-connected provided that for any point p of M there exists an arbitrarily small neighborhood of p in M whose complement has only a finite number of components. See my paper in the *American Journal of Mathematics*, vol. 61 (1939), pp. 733-749.

THEOREM 4. *If A is a compact semi-locally connected continuum, a necessary and sufficient condition that every irreducible transformation on A be strongly irreducible is that the non-cut points of A be dense in A.*

The sufficiency of the condition follows at once from Theorems 2 and 3. To see that the condition is necessary we need only note that if the non-cut points of a continuum A are not dense in A, then A contains a free arc, say $axuvyb$, of cut points; and if we fold this arc so that v goes into x, u into y, and xu, vu and vy go into an arc xy and all other points of A go into themselves, we set up a transformation on A which is irreducible but not strongly irreducible.

THEOREM 5. *If A is a semi-locally connected compact continuum, any irreducible interior transformation $T(A) = B$ is a homeomorphism.*

Proof. Suppose, on the contrary, that for some $b \in B$, $T^{-1}(b) \supset x + y$ where $x \neq y$. Then x must be a cut point of A. For if not, let V_x and V_y be disjoint neighborhoods of x and y respectively so chosen that $T(V_x) \subset T(V_y)$; there exists [3] a continuum N such that $A - V_x \subset N \subset A - x$; whence $T(N) \supset T(V_y) \supset T(V_x)$ so that $T(N) = T(A)$, contrary to the irreducibility of T. Now let R be the component of $A - x$ containing y. Then R must map onto all of B. For let Q be any component of $B - b$. Since [3] R is open there exists a component W of $T^{-1}(Q)$ with $W \cdot R \neq 0$; and since $W \cdot x = 0$, it follows that $W \subset R$. Hence $T(R) \supset Q$. Thus $T(R) \supset B$. But $R + x$ is a proper subcontinuum of A. Hence we have a contradiction and our theorem is proven.

UNIVERSITY OF VIRGINIA.

PROPERTY S_n.* †

By R. L. Wilder.

The work of Sierpinski [1] and Moore [2] on the implications of Property S, especially in characterizing continuous curves (Peano continua) from both intrinsic and positional points of view, as well as the numerous uses of the property by other authors in the study of locally connected spaces,[3] suggests that the notion should be generalized and relations to higher dimensional local connectedness established.

As defined by Moore,[2] Property S requires that the point set in question be, for each $\epsilon > 0$, the sum of a finite number of connected sets each of which is of diameter less than ϵ. He found that this property is weaker than that of uniform local connectedness and stronger than that of local connectedness.

In the present paper we define a property which we call Property S_n, which for $n = 0$ is equivalent to the Sierpinski-Moore Property S, and insofar as the present investigation goes is shown to yield the "justification theorems" which a generalization of Property S might be expected to yield. For example, the Sierpinski characterization of Peano continua is the case $n = 0$ of a characterization of the general lc^n in terms of S-properties, as also are the relations mentioned above as established by Moore between the $ul0 - c.$, S_0 and $l0 - c.$ properties. We also consider positional properties of sets imbedded in n-space; for example, we find a duality between the lc^k property and the S-properties of the complement.

I. Definitions.

By a *pair* (U, V) we mean an ordered pair of point sets U and V such that $U \supset V$. By $h^i(U, V)$ we denote the maximum number of i-cycles [4] of V that are linearly independent with respect to bounding on U. A set of pairs

* Received May 8, 1939.

† Presented to the American Mathematical Society, September 7, 1937.

[1] W. Sierpinski, "Sur une condition pour qu'un continu soit une courbe jordanienne," *Fundamenta Mathematicae*, vol. 1 (1920), pp. 44-60.

[2] R. L. Moore, "Concerning connectedness im kleinen and a related property," *ibid.*, vol. 3 (1922), pp. 232-237; also his book, *Foundations of Point Set Theory*, pp. 271-280.

[3] Particularly G. T. Whyburn; for example, see his papers, "A note on spaces having the S property," *American Journal of Mathematics*, vol. 54 (1932), pp. 536-538; "Concerning S-regions in locally connected continua," *Fundamenta Mathematicae*, vol. 20 (1933), pp. 131-139. See also R. L. Wilder, "Concerning perfect continuous curves," *Proceedings of the National Academy of Sciences*, vol. 16 (1930), pp. 233-240.

[4] In the sense of Vietoris.

(U, V) will be said to *cover* a point set M if for each $p \varepsilon M$ there exists at least one pair (U, V) of the given set such that $p \varepsilon V$.

If P is a set of pairs (U, V), by $\Sigma_U P$ we mean the set-theoretic sum of all U's involved in elements of P; and by $\Sigma_V P$ we mean the set-theoretic sum of the V's involved in elements of P. The number $h^i(\Sigma_U P, \Sigma_V P)$ we abbreviate to $h^i(P)$. A point set M will be called the *sum* of a set P of pairs (U, V) if $M = \Sigma_U P = \Sigma_V P$.

By the diameter $\delta(U, V)$ of a pair (U, V) we shall mean the diameter of $U, \delta(U)$.

A point set M will be said to have *property* S_n if for each $\epsilon > 0$ it is the sum of a finite set P of pairs of diameters less than ϵ, such that for every subset P' of P the number $h^n(P')$ is finite. Clearly *a set cannot have property* S_n *unless its n-dimensional Betti number is finite.*

This property may be specialized in various ways; for example, if B^n represents any special group of n-cycles of M, we may say that M has *property* S_n *relative to the group* B^n if it satisfies the above requirement for S_n but with the stipulation that in the previous definition of $h^i(U, V)$ the words "maximum number of i-cycles" be replaced by "maximum number of cycles of B^i."

THEOREM 1. *Properties* S and S_0 *are equivalent.*

Proof. Suppose M has property S, and hence for any given $\epsilon > 0$ is the sum of a finite number of connected sets M_i such that $\delta(M_i) < \epsilon$. For each i, let $(U, V)_i$ denote a pair in which $U = V = M_i$, and let P be the set of all pairs $(U, V)_i$. Since for any subset P' of P the number of components of $\Sigma_U P'$ is finite, the number $h^0(P')$ is finite, and therefore M has property S_0.

Conversely, suppose M has property S_0, and let $\epsilon > 0$ be given. Then M is the sum of a finite set P of pairs $(U, V)_i$ such that $\delta(U, V)_i < \epsilon$ and for every $P' \subset P$, $H^0(P')$ is finite. In particular, each $h^0(U, V)_i$ is finite, implying that V lies in a finite number m_i of components of U. Since $M = \Sigma_V P$, it follows that M is the sum of at most $\sum_{i=1}^{i=k} m_i$ connected sets of diameters less than ϵ (where k is the number of elements in P).

II. Relations to local connectedness.

THEOREM 2. *Property* S_n *is stronger than local n-connectedness,*[5] *even (when* $n > 0$*) for compact spaces.*

[5] For the case $n = 0$, this was proved by Moore, *loc. cit.*, Theorem 2.

Proof. Suppose M has property S_n and $\epsilon > 0$ is given. Then M is the sum of a finite set P of pairs (U, V) such that $\delta(U, V) < \epsilon/4$ and $h^n(P')$ is finite for all $P' \subset P$. For any $x \, \varepsilon \, M$, let P' be the set of all those pairs (U, V) which contain points of the spherical neighborhood $S(x, \epsilon/4)$. Then $S(x, \epsilon/4) \subset \Sigma_V P' \subset \Sigma_U P' \subset S(x, \epsilon)$. At most a finite number of n-cycles of $S(x, \epsilon/4)$ are independent with respect to bounding in $S(x, \epsilon)$ and hence M is locally n-connected at x.[6]

That a non-compact set may be locally 0-connected and not have property S^0 follows from simple examples;[2] for compact sets the two properties are equivalent.[1] For $n > 0$, however, a compact set may be locally n-connected yet not have property S_n; thus for $n = 1$ we may construct an example as follows:

Example 1. In the coördinate plane let p_n denote the point $(1/n, 0)$, $n = 1, 2, 3, \cdots$. Let K_n be the set of points on an ellipse with center p_n, major axis parallel to the y-axis and length 2, and minor axis of length less than $[\rho(p_n, p_{n+1})]/2$. Let K_0 denote the set of all points $(0, y)$ such that $-1 \leqq y \leqq 1$. Then the point set $M = \sum_{n=0}^{\infty} K_n$ is compact and locally 1-connected, but does not have property S_1. (This is evident since the Betti number $p^1(M)$ is infinite.)

As another example, consider the following:

Example 2. In coördinate 3-space let K_n be the set of all points (x, y, z) such that $0 \leqq x \leqq 1$, $y = 1/n$ for $n = 1, 2, 3, \cdots$ and $0 \leqq z \leqq 1$. Let K_0 be the surface, excepting the base in the xy-plane, of the unit cube of whose faces three lie in the coördinate planes and of which $(1, 1, 1)$ is a vertex. Let $M = \sum_{n=0}^{\infty} K_n$. Here again M is compact and locally 1-connected, and does not have property S_1. In this case, however, $p^1(M) = 0$. If P were a set of pairs of diameters less than $\frac{1}{2}$, whose sum is M, and P' were that subset of P consisting only of those pairs containing points for which $z = 0$, then $h^1(P')$ would be infinite.

We shall see, however, that the situation changes whenever a set has property S_n for all values of n up to and including some positive integer. For sake of brevity, we say that a set has property S_j^k if it has property S_i for all i such that $j \leqq i \leqq k$. By lc^n we denote a compact metric space which is locally i-connected for all $i \leqq n$. We give first the generalization of the Sierpinski characterization of Peano continua.

---

[6] See P. Alexandroff, "On local properties of closed sets," *Annals of Mathematics*, vol. 36 (1935), pp. 1-35 (especially p. 9).

THEOREM 3. *In order that a compact metric space should be an lc^n, it is necessary and sufficient that it have property $S_0{}^n$.*

Proof. Let M be an lc^n, and let i be any integer such that $0 \leqq i \leqq n$. Let $\epsilon > 0$ be arbitrary. If $x \,\varepsilon\, M$, there exists $\delta_x > 0$ such that all i-cycles of $S(x, \delta_x)$ bound on $S(x, \epsilon)$. Since M is compact, a finite number of the open sets $S(x, \delta_x)$ covers M. Considering the latter as sets V, and associating with each such V the corresponding $U = S(x, \epsilon)$ we obtain a finite set P of pairs (U, V) whose sum is M.

Let P' be any subset of P. Then the closure of the set $\Sigma_V P'$ is a subset of the open set $\Sigma_U P'$, and consequently the number $h^i(P')$ is finite.[7] Thus M has property S_i.

That the condition of the theorem is sufficient follows from Theorem 2.

III. Application to positional properties of subsets of E_n.

In his paper cited above,[2] Moore showed that in order that a simply connected bounded domain D in the plane should have a peanian boundary it is necessary and sufficient that D have property S. In this section we consider analogous theorems for the general euclidean space E_n.

THEOREM 4. *In E_n let D be a simply $(n-1)$-connected[8] domain which has property $S_0{}^{n-2}$. Then the boundary B of D is a Peano continuum.*

Proof. Consider the set $M = E_n - D$. As D is simply $(n-1)$-connected, M is a continuum. If we can show that M is an lc^{n-2}, it will follow from Theorem 7 of L. C. that B is a Peano continuum.

Let R and W be open subsets of E_n such that $R \supset \bar{W}$. Then we shall show for any i such that $0 \leqq i \leqq n-2$, that $h^i(R-M, W-M)$ is finite, from which it will follow that M is locally i-connected.[9] Let us select a positive number ϵ less than $\rho[\bar{W}, F(R)]/2$. By hypothesis, D is the sum of a finite set P of pairs (U, V) of diameters less than ϵ, such that for any $P' \subset P$, $h^i(P')$ is finite. Let P' be the set of all those elements (U, V) of P such that

---

[7] By Theorem 2 of my paper, "On locally connected spaces," *Duke Mathematical Journal*, vol. 1 (1935), pp. 543-555; this paper will be referred to hereafter as L. C. A finite coefficient group for chains and cycles is assumed here.

[8] I. e., $p^{n-1}(D) = 0$.

[9] In my (unpublished) paper "Locally connected subsets of euclidean n-space," (see *Bulletin of the American Mathematical Society*, vol. 42 (1936), p. 496, abstract no. 308), it is shown that a necessary and sufficient condition that a compact closed subset M of E_n be an lc^k is that if R and W are as defined above, then $h^{n-i-1}(R-M, W-M)$ is finite for $0 \leqq i \leqq k$, and of the bounding $(n-k-2)$-cycles of $W-M$, at most a finite number are linearly independent with respect to bounding in $R-M$.

V contains a point of W. The set P' covers $W \cdot D = W - M$, and $\Sigma_U P' \subset R \cdot D = R - M$. Since $h^i(P')$ is finite, it follows that $h^i(R - M, W - M)$ is finite.

Simple examples may be constructed in E_3 to show that the converse of Theorem 4 does not in general hold.

THEOREM 5. *In order that a continuum M in H_n*[10] *should be peanian, it is necessary and sufficient that its complement have property S_{n-2} relative to the group B^{n-2} of $(n-2)$-cycles that bound in $H_n - M$.*

Proof. The necessity may be proved as follows: Given $\epsilon > 0$ and $x \varepsilon H_n - M$, let $U^* = S(x, \epsilon)$ and $V^* = S(x, \epsilon/2)$. Let P^* be a finite set of pairs (U^*, V^*) covering $H^n - M$. The set P of all pairs (U, V) such that $U = U^* \cdot (H_n - M)$, $V = V^* \cdot (H_n - M)$, $(U^*, V^*) \varepsilon P^*$, is a set of pairs whose sum is $H_n - M$. Consider any $P' \subset P$. As shown in the paper cited above,[9] at most a finite number of elements of B^{n-2} in $\Sigma_V P'$ are independent with respect to bounding in $\Sigma_U P'$. Consequently $H^n - M$ has property S_{n-2} relative to B^{n-2}.

Proof of sufficiency. Let $H^n - M$ have property S_{n-2} relative to B^{n-2}, and suppose M is not peanian. Then there exist $x \varepsilon M$ and numbers $\epsilon > \delta > \eta > 0$ such that of the cycles of B^{n-2} that lie on $(H_n - M) \cdot F(x, \delta)$, infinitely many are linearly independent with respect to bounding in $(H_n - M) \cdot [S(x, \epsilon) - \bar{S}(x, \eta)]$. Let θ be a positive number less than $\frac{1}{2}$ max $(\epsilon - \delta, \delta - \eta)$.

By hypothesis, $H_n - M$ is the sum of a set P of pairs (U, V) of diameter $< \theta$ such that for any $P' \subset P$, at most a finite number of cycles of B^{n-2} in $\Sigma_V P'$ are independent with respect to bounding in $\Sigma_U P'$. But consider the set P' of all elements (U, V) of P such that $V \cdot F(x, \delta) \neq 0$. The set P' covers $(H_n - M) \cdot F(x, \delta)$, and consequently only finitely many elements of B^{n-2} on $F(x, \delta)$ are independent with respect to bounding in $\Sigma_U P'$. But $\Sigma_U P' \subset (H_n - M) \cdot [S(x, \epsilon) - \bar{S}(x, \eta)]$ and a contradiction results.

For the case of the general lc^k we state the following characterization in terms of positional properties:

THEOREM 6. *In order that a closed subset M of H_n should be an lc^k it is necessary and sufficient that its complement have property S^{n-1}_{n-k-1} as well as property S_{n-k-2} relative to the group B^{n-k-2} of $(n-k-2)$-cycles that bound in the complement.*

Proof. The proof of the necessity is analogous to the proof of the necessity in Theorem 5. For the sufficiency, it suffices to prove[9] that if R and W

[10] By H_n we mean the euclidean n-sphere.

are open subsets of H_n such that $R \supset \bar{W}$, then at most a finite number of elements of B^{n-k-2} that lie in W are independent with respect to bounding in $R - M$ and a similar statement holds for the groups Z^{n-i-1}, $i = 0, 1, \cdots, k$, of $(n - i - 1)$-cycles of $H_n - M$. That this is the case may be proved in a manner analogous to that of the second paragraph of the proof of Theorem 4.

The case $k = n - 2$ of Theorem 6 is of special interest, inasmuch as the sets lc^{n-2} seem to form, in H_n, the natural analogue of the peanian continua in the plane. In Theorem 6, when $k = n - 2$, the requirement that the complement of M have property S_0 for the bounding 0-cycles is imposed. This implies not only that the domains complementary to M have property S_0, but that their diameters form a null sequence.[11] As a matter of fact, the following theorem holds:

THEOREM 7. *In order that the complement of a closed subset M of H_n should have property S_0 relative to the group B^0 of bounding 0-cycles of the complement, it is necessary and sufficient that the domains of $H_n - M$ all have property S_0 and their diameters form a null sequence.*

Proof. For the sufficiency we may argue as follows: For fixed $\epsilon > 0$, and each $x \in H_n - M$, let $U' = S(x, \epsilon)$ and $V' = S(x, \epsilon/2)$. Let P' be a finite number of pairs (U', V') covering the closure of $H^n - M$. Then it easily follows that all but a finite number of the domains complementary to M have the property that each is covered by some element of P'; those that do not have this property we denote by $D_i, i = 1, \cdots, k$. Let $L = \sum_{i=1}^{k} D_i$ and for $(U', V') \in P'$ let $U = U' \cdot (H_n - M - L)$, $V = V' \cdot (H_n - M - L)$; let P_1 denote the set of corresponding pairs (U, V). Since each D_i has property S_0, there exists a set P_2 of pairs of diameters $< \epsilon$ whose sum is L, and such that for any $P_2' \subset P_2$ the number of cycles of B^0 in $\Sigma_V P_2'$ independent with respect to bounding in $\Sigma_U P_2'$ is finite. The set of pairs $P = P_1 + P_2$ is easily seen to have the desired properties.

The necessity is the result of a more general theorem, namely,

If a metric space A has property S_0 relative to its bounding 0-cycles, then its non-degenerate components are at most denumerable in number, each has property S_0, and their diameters form a null sequence.

Proof. That each component C will have property S_0 is easily shown by obtaining, for $\epsilon > 0$, the desired set P for A and for each $(U, V) \in P$ retaining the pair $(U \cdot C, V \cdot C)$—the 0-cycles of C being bounding 0-cycles of M.

---

[11] Compare R. L. Wilder, "Sets which satisfy certain avoidability conditions," *Časopis pro Pěstování Matematiky a Fysiky*, 1937/38, pp. 185-198, Theorem 5.

Suppose A has an infinite number of components C_i of diameter greater than some positive number ϵ, $i = 1, 2, 3, \cdots$. Since A has property S_0 relative to its bounding 0-cycles, there is a finite set P of pairs of diameter $< \epsilon/4$ whose sum is A and such that only a finite number of the bounding 0-cycles of A in any $\Sigma_V P'$ are independent with respect to bounding in the corresponding $\Sigma_U P'$. But it is readily shown that in each C_i there exist points p_i and q_i such that $\rho(p_i, q_i) \geqq \epsilon$, and elements (U_1, V_1), (U_2, V_2) of P such that infinitely many of the points p_i lie in V_1 and infinitely many of the points q_i lie in V_2; from this situation a contradiction results.

The following theorem is a corollary of Theorem 6 and Principal Theorem B of the paper cited in footnote [11].

THEOREM 8. *If M is a simply i-connected $(i = 0, 1, \cdots, n-2)$ subcontinuum of H_n, which has no i-cut-points and whose complement has property S_0 relative to its bounding 0-cycles and property S_1^{n-2}, then the boundaries of the domains complementary to M are all simply i-connected generalized closed $(n-1)$-manifolds; in particular, if $n = 3$, these boundaries are all 2-spheres.*

If we do not require, in Theorem 8, the simple i-connectedness and replace absence of cut-points by local avoidability, we may state:

THEOREM 9. *If the complement of a locally i-avoidable $(i = 0, 1, \cdots, n-2)$ subcontinuum of H_n has property S_0 relative to its bounding 0-cycles and property S_1^{n-2}, then all but a finite number of the complementary domains of M are bounded by generalized closed $(n-1)$-manifolds.*

Theorem 9 is a corollary of Theorem 6 above and Theorem 10 of the paper cited in footnote [11].

IV. Relations to uniform local connectedness.

We shall conclude with some results concerning the relations between *uniform local n-connectedness* [12] $(= uln - c)$ and property S_n. That for an isolated $n > 0$ a set may be $uln - c$. and not have property S_n is shown by Example 1 above. As for the converse, consider the set B in E_3 consisting of a spherical surface and one of its radii. If D is the bounded domain complementary to B, then D has properties S_0 and S_1 by Theorems 6 and 7, but is not $ul1 - c$. We shall find, however, that if a set is a ulc^n, that is, $uli - c$, for all $i \leqq n$, then it does have property S_0^n.

---

[12] A metric space is $uln - c$, if for arbitrary $\epsilon > 0$ there exists a $\delta > 0$ such that every n-cycle of diameter $< \delta$ bounds on a self-compact set of diameter $< \epsilon$ (only cycles with self-compact carriers are considered).

LEMMA 1. *If M is a ulc^n, then for any $\epsilon > 0$ there exists a $\delta > 0$ such that any abstract λ-dimensional $(0 \leqq \lambda \leqq n+1)$ cell of M of diameter $< \delta$ has a chain-realization in M of diameter $< \epsilon$.*

LEMMA 2. *If a subset M of a metric space is a ulc^n, then for any $\epsilon > 0$ there exists a $\delta > 0$ such that any δ-chain of dimension λ $(0 \leqq \lambda \leqq n+1)$ of $\bar{M}$ has an ϵ-removed chain-realization in M.*[13]

THEOREM 10. *If a subset M of a metric space is a ulc^n, then $\bar{M}$ is an lc^n.*[14]

Proof. Given $\epsilon > 0$ and $p \varepsilon \bar{M}$, let $\delta > 0$ be such that any i-cycle of M $(0 \leqq i \leqq n)$ of diameter $< \delta$ bounds a chain of M of diameter $< \epsilon/4$. Let γ^i be a cycle of $\bar{M} \cdot S(p, \delta)$. It need only be shown that for any $\eta > 0$ there exists a number N such that if $\gamma^i = \{i_m\}$ and $m > N$, then $i_m \underset{\eta}{\sim} 0$ in $\bar{M} \cdot S(p, \epsilon)$. By applying Lemma 2, it is readily shown that for m large enough the chains i_m have η-removed chain realizations in $M \cdot S(p, \delta)$ which bound chains of M which are of diameter $< \epsilon/4$ and hence lie in $M \cdot S(p, \epsilon)$. Consequently there exists a value of N such as that described above.

LEMMA 3. *Let M be a ulc^n. Then every i-cycle $(0 \leqq i \leqq n)$ $\gamma^i = \{i_m\}$ of M is homologous on an arbitrary neighborhood of its carrier to chain-realizations of all the cycles i_m for m sufficiently large.*[15]

Proof. Let $\gamma^i = \{i_m\}$ be an i-cycle of M with carrier F_0.[16] Since M is a ulc^n, the cycles i_m, for m large enough, have chain-realizations $z_m{}^i$ and the homologies $i_m \underset{\epsilon_m}{\sim} i_{m+1}$, where $\lim \epsilon_m = 0$, have chain-realizations $K_m{}^{i+1}$, where $K_m{}^{i+1} \rightarrow z_m{}^i - z_{m+1}{}^i$, carried by self-compact subsets F_m of M such that the sets F_m converge uniformly to F_0.

Given $\epsilon > 0$, we choose m large enough so that for all $k \geqq m$, $F_k \subset S(F, \epsilon)$. Let $Q = \sum_{k=m}^{\infty} F_k$. Then, considering m fixed from now on, $\gamma^i \sim z_m{}^i$ on Q. Since Q is self-compact, we need only show that for arbitrary $\eta > 0$, $z_m{}^i \underset{\eta}{\sim} \gamma^i$ on Q.

Let $z_m{}^i = \{z_{ms}{}^i\}$. We fix $k \geqq m$ so that $i_k \underset{\eta}{\sim} i_s$ for all $s \geqq k$ on F_0; $i_k \underset{\eta}{\sim} z_k{}^i$ on F_k and hence $i_k \underset{\eta}{\sim} z_{ks}{}^i$ on F_k for all $s \geqq k$.

---

[13] For definitions, and indications of proof of these lemmas, see L. C., especially Lemma 1; also Lemma 2 of my paper "Generalized closed manifolds in n-space," *Annals of Mathematics*, vol. 35 (1934), pp. 876-903. Whereas in the latter case the ϵ-removed realizations were actually polyhedral, the set in question being an open subset of n-space, in the present instance the ϵ-removed realizations are chain-realizations as defined in L. C.

[14] For the case $n = 0$, this theorem was proved by Moore.[2]

[15] Compare with P. Alexandroff, "Zur Homologie-Theorie der Kompakten," *Compositio Mathematica*, vol. 4 (1937), pp. 256-270, Satz II of § 3.

[16] By a carrier of γ^i we mean a self-compact set F_0 which not only carries the cycles i_m but also the homologies relating them.

Since $z_m{}^i \sim z_k{}^i$ on Q, there exists N such that for $s \geqq N$ and $s \geqq k$, $z_{m_s}{}^i \,\bar{\eta}\, z_{k_s}{}^i$ on Q. Then for such values of s, we have $i_s \,\bar{\eta}\, i_k \,\bar{\eta}\, z_{k_s}{}^i \,\bar{\eta}\, z_{m_s}{}^i$ on Q.

THEOREM 11. *If M is a ulc^n such that $\bar{M}$ is compact, then M has property $S_0{}^n$.*

Proof. Given $\epsilon > 0$, for each $p \,\varepsilon\, \bar{M}$ let $U' = \bar{M} \cdot S(p, \epsilon)$ and $V' = \bar{M} \cdot S(p, \epsilon/2)$. As $\bar{M}$ is compact, a finite set P' of such pairs (U', V') covers $\bar{M}$. For each $(U', V') \,\varepsilon\, P'$ let $U = M \cdot U'$, $V = M \cdot V'$, and let P be the set of resulting pairs (U, V). Then M is the sum of the pairs in P. Let P_1 be any subset of P. Let P'_1 be the set of elements of P' corresponding to the elements of P_1.

Since by Theorem 10, $\bar{M}$ is an lc^n, it follows from Theorem 2 of L. C. that at most a finite number, say m, of i-cycles $(0 \leqq i \leqq n)$ of $\Sigma_V P'_1$ are independent with respect to bounding in $\Sigma_U P'_1$. Let $\gamma_h{}^i$, $h = 1, 2, \cdots, m+1$ be a set of i-cycles of $\Sigma_V P_1$. Then since these are also cycles of $\Sigma_V P'_1$, there is a bounding relation $K^{i+1} \to \sum_{h=1}^{m+1} c^h \gamma_h{}^i$, where K^{i+1} is a chain of $\Sigma_U P'_1$. For the sake of brevity let us denote the cycle $\sum_{h=1}^{m+1} c^h \gamma_h{}^i$ by Γ^i. The carrier of Γ^i in M is a self-compact set F.

Let $K^{i+1} = \{k_m{}^{i+1}\}$ and $\Gamma^i = \{i_m\}$, where $k_m{}^{i+1} \to i_m$ and $k_m{}^{i+1}$ is an ϵ_m-chain, $\lim \epsilon_m = 0$. By Lemma 3 there exists a number N such that for $m \geqq N$, $\Gamma^i \sim z_m{}^i$ on $\Sigma_V P_1$, where $z_m{}^i$ is a chain-realization of i_m. By application of Lemma 2 there exists a number N' such that for $m \geqq N'$, the chain $k_m{}^{i+1}$ has an η_m-removed realization $L_m{}^{i+1}$ in $\Sigma_U P_1$, $\lim \eta_m = 0$. The realization of $L_m{}^{i+1}$ can be carried out in such a manner that $L_m{}^{i+1} \to z_m{}^i$. In particular, for $m > \max(N, N')$, we have $L_m{}^{i+1} \to z_m{}^i$ on $\Sigma_U P_1$. It follows that Γ^i bounds on $\Sigma_U P_1$ and therefore that $h^i(P_1)$ is finite.

V. Appendix.

In Theorems 8 and 9 we have seen that certain avoidability properties of a continuum in H_n combined with suitable S-properties of the complement will yield general manifold boundaries for the complementary domains. Since this paper was written, there has appeared a paper of G. T. Whyburn [17] in which the S_0-property of domains complementary to a plane continuum M is a result of an avoidability property of M. For reasons which will become apparent, we reformulate Whyburn's result as follows:

We call a metric space M *almost i-avoidable* at $p \,\varepsilon\, M$ if for $\epsilon > 0$ there exists $\delta > 0$ such that at most a finite number of i-cycles of $M \cdot F(p, \epsilon)$ are

[17] "Semi-locally connected sets," *American Journal of Mathematics*, vol. 61 (1939), pp. 733-749.

linearly independent with respect to bounding on $M - S(p, \delta)$.[18] Whyburn's result[19] may be stated: *If M is an almost 0-avoidable continuum in H_2, then each complementary domain of M has property S_0.* We may obtain this result as the case $n = 0$ of the following theorem:

THEOREM. *Let M be an almost $(n-2)$-avoidable, closed subset of the euclidean n-sphere H_n. Then each complementary domain of M has property S_0.*

Proof. Suppose D, a domain complementary to M, does not have property S_0. Then there exist $p \varepsilon M$, $\epsilon > 0$, and a sequence of points $p_i \varepsilon D$ converging to p and such that no two points p_i can be joined by an arc of $D \cdot S(p, \epsilon)$. By hypothesis there exists $\delta > 0$ such that on $M \cdot F(p, \epsilon)$ at most a finite number, say m, of $(n-2)$-cycles are independent with respect to bounding on $M - S(p, \delta)$.

By local duality properties[20] there exist on M infinitely many cycles $\gamma_k^{n-1} \bmod M - S(p, \epsilon)$ that are uniquely linked in $S(p, \epsilon)$ with cycles γ_k^0 of $D \cdot S(p, \delta)$ based on pairs of the points p_i. The cycles γ_k^{n-1} cannot be absolute, since the p_i's are all in one component of $H_n - M$. Thus, as a chain, each γ_k^{n-1} is bounded by an absolute γ_k^{n-2} of $F(p, \epsilon)$. By the avoidability assumption there exists a bounding relation $K^{n-1} \rightarrow \sum_{k=1}^{m+1} c^k \gamma_k^{n-2}$ on $M - S(p, \delta)$. Denoting $\sum_{k=1}^{m+1} c^k \gamma_k^{n-1}$ by γ^{n-1}, the chain $\Gamma^{n-1} = \gamma^{n-1} - K^{n-1}$ is an absolute cycle of M.

At least one of the c^k's, say c^1, is not zero. Then the cycle γ^{n-1} is linked with $c^1\gamma_1^0$ in $S(p, \epsilon)$; in particular, if $L^1 \rightarrow c^1\gamma_1^0$ in $S(p, \delta)$, the intersection number $\chi(\gamma^{n-1}, L^1) \neq 0$. But there exists $K^1 \rightarrow c^1\gamma_1^0$ in D for which $\chi(\Gamma^{n-1}, K^1) = 0$. By the invariance of the intersection number we have $\chi(\Gamma^{n-1}, K^1) = \chi(\Gamma^{n-1}, L^1) = \chi(\gamma^{n-1}, L^1) = 0$, thus giving a contradiction.

The same type of argument also gives the theorem, *if M is almost i-avoidable, $p \varepsilon M$ and $\epsilon > 0$, then there exists $\delta > 0$ such that at most a finite number of the $(n-i-2)$-cycles of $(H_n - M) \cdot S(p, \delta)$ that bound in $H_n - M$ are independent with respect to bounding in $(H_n - M) \cdot S(p, \epsilon)$.*

UNIVERSITY OF MICHIGAN.

---

[18] See my paper referred to above,[11] Definition III. Definitions I and II of the same paper may likewise be modified to yield corresponding "almost avoidable" properties—in this sense, the property of being almost completely avoidable is characteristic of lc^n sets. It will be observed that Whyburn's "semi-local connectedness" and the property "almost 0-avoidable" are equivalent for connected spaces.

[19] The theorem of § III. As stated by Whyburn, the conclusion of this theorem reads, "the boundary of each complementary domain of M is locally connected"; however, his method of proof is to obtain this as a corollary of the S-property of the complementary domains.[2]

[20] See P. Alexandroff, "On local properties of closed sets," *loc. cit.*, section 2.

ON CIRCULATION FUNCTIONS.* [1]

By VINCENT C. POOR.

1. Introduction. In an article [2] by Priwaloff, the Cauchy singular integral is studied. Under certain integrability conditions on the function involved he proves that if the *principal value* of the singular integral exists the other limits exist and conversely; these other limits are expressed in terms of the principal value.

In this note the existence of the principal value for a class of polygenic functions, called circulation functions, will be proved. The necessary and sufficient condition for the existence of such functions will then be obtained; this condition involves the principal value of the Cauchy singular integral for the functions studied. Incidentally, an extension of the Pompeiu Theorem will be required.

2. Circulation functions defined. In a region of the complex plane let c be a simple rectifiable curve bounding the area σ; then the limit,

$$\lim_{\sigma\to 0} \frac{\int_c f(z)\,dz}{2i\sigma} = \frac{\partial f}{\partial \alpha} \tag{2.1}$$

is by definition the areal [3] derivative when the limit exists. The notation here used appeared in a previous article.[4] By dividing the original area σ into cells by a system of mesh circuits one easily argues the integral form of (2.1) given by

$$\frac{1}{2\pi i}\int_c f(z)\,dz = \frac{1}{\pi}\int_\sigma \frac{\partial f}{\partial \alpha}\,d\sigma. \tag{2.2}$$

All contour integrals will be taken in the positive sense, that is, with the area to the left. The circulation theorem (2.2), valid for all rectifiable closed curves in a region D, defines a class of polygenic functions in D. Every such function is then by definition a *circulation function* in D.

* Received November 22, 1937; Revised October 31, 1938.

[1] Presented to the Society at Indianapolis, December, 1937.

[2] J. Priwaloff, "Sur certaines propriétés métriques des fonctions analytiques," *Jr. de l'Ecole Polytechnique* (1924), p. 96.

[3] D. Pompeiu, *Rendiconti di Palermo*, vol. 35 (1913), p. 278.

[4] V. C. Poor, *Transactions of the American Mathematical Society*, vol. 32, no. 2, p. 216.

If u is a point outside the fixed contour c, (2.2) may be written in the form

$$(2.3)\qquad \frac{1}{2\pi i}\int_c \frac{f(z)\,dz}{z-u} = \frac{1}{\pi}\int_\sigma \frac{\frac{\partial f}{\partial \bar{z}}\,d\sigma}{z-u}.$$

If u is a point of σ an application of (2.3) to a continuous circulation function $f(z)$ leads to the Pompeiu theorem,[5]

$$(2.4)\qquad \frac{1}{2\pi i}\int_c \frac{f(z)\,dz}{z-u} = \frac{1}{\pi}\int_\sigma \frac{\frac{\partial f}{\partial \bar{z}}\,d\sigma}{z-u} + f(u)$$

which will be useful in the sequel.

3. The singular integral. Assume that the contour c bounding σ has a tangent almost everywhere. The Cauchy integral

$$\frac{1}{2\pi i}\int_c \frac{f(z)\,dz}{z-u}$$

becomes the Cauchy singular integral when u becomes a point x of the curve.

It is singular in the sense that it has no meaning; yet three distinct limiting values may be defined. If x_1 and x_2 are points on c near to x so chosen that the arcs x_2x and xx_1 are equal and if c_1 is the remaining segment x_1x_2 which does not contain x then

$$\operatorname*{limit}_{x_2x=xx_1\to 0} \frac{1}{2\pi i}\int_{c_1} \frac{f(z)\,dz}{z-x}$$

is called the principal value of the singular integral when this limit exists. The symbol

$$\frac{1}{2\pi i}\int_c \frac{f(z)\,dz}{z-x}$$

will be written for this principal value; ϕ_i and ϕ_e will designate the other limit values obtained by making u approach x from the inside and outside respectively.

In § 4 a different definition is used in determining the principal value; the two arcs in this definition, however, approach equality.

The Priwaloff results, when the principal value exists, are the following:

$$(3.1)\qquad \begin{aligned} \phi_i &= \frac{1}{2\pi i}\int_c \frac{f(z)\,dz}{z-x} + \tfrac{1}{2}f(x) \\ \phi_e &= \frac{1}{2\pi i}\int_c \frac{f(z)\,dz}{z-x} - \tfrac{1}{2}f(x). \end{aligned}$$

[5] D. Pompeiu, *loc. cit.*

4. The existence of the principal value. It will now be possible to prove the existence of the principal value of the singular integral for the class of continuous circulation functions defined in D.

To do this assume that the boundary c lying entirely in D has a unique tangent almost everywhere; then with x on c as a center describe a small circle, o, radius r intersecting c. It is assumed that x is so chosen that the curve has a unique tangent at the point. Let c_1 be the part of c outside o, and let o_1 and o_2 be the arcs of o inside and outside c respectively; then for the area bounded by $c_1 + o_2$, (2.4) is valid when u is replaced by x since x is a point of the area, or

$$(4.1) \qquad \frac{1}{2\pi i}\int_{c_1+o_2} \frac{f(z)\,dz}{z-x} = \frac{1}{\pi}\int_{\sigma+\sigma_2} \frac{\dfrac{\partial f}{\partial \alpha}\,d\sigma}{z-x} + f(x)$$

where σ_2 is the area between c and o_2. For the area bounded by $c_1 + o_1$, since x is outside the area, (2.3) is applicable so that

$$(4.2) \qquad \frac{1}{2\pi i}\int_{c_1+o_2} \frac{f(z)\,dz}{z-x} = \frac{1}{\pi}\int_{\sigma-\sigma_1} \frac{\dfrac{\partial f}{\partial \alpha}\,d\sigma}{z-x}$$

where σ_1 indicates the area between c and o_1.

The integral along o_2 is

$$\int_{o_2} \frac{f(z)\,dz}{z-x} = f(x)\int_{o_2} \frac{dz}{z-x} + \int_{o_2} \frac{f(z)-f(x)}{z-x}\,dz$$

The second integral on the right goes to zero with r while

$$f(x)\int_{o_2} \frac{dz}{z-x} = if(x)\int_0^{\pi+\theta_1} d\theta = i(\pi+\theta_1)f(x)$$

where θ_1 vanishes with r. The corresponding integral along o_1 is $(-\pi+\theta_1) \times f(x)$.

When (4.1) and (4.2) are added and r made to approach zero the sum of the integrals along o_1 and o_2 will vanish with r; σ_1 and σ_2 will each approach zero; in the limit one finds that

$$(4.3) \qquad \frac{1}{2\pi i}\int \frac{f(z)\,dz}{z-x} = \frac{1}{\pi}\int_{\sigma} \frac{\dfrac{\partial f}{\partial \alpha}\,d\sigma}{z-x} + \tfrac{1}{2}f(x).$$

Thus the principal value of the singular integral exists and it is given by (4.3). The other limit values as given in (3.1) may be obtained from (2.3) and (2.4) by making u approach x properly.

This result may also be stated as

THEOREM 1. *If $f(z)$ is a continuous circulation function on σ it necessarily follows that $f(z)$ satisfies the functional equation* (4. 3).

5. Existence of circulation functions. As is well known, if $f(z)$ is monogenic on σ, it is expressible as a Cauchy integral. However, the integral

$$\frac{1}{2\pi i}\int_c \frac{f(z)\,dz}{z-u}$$

defines a monogenic function on σ if and only if $f(z)$ on the boundary is suitably chosen.

Similarly, the Pompeiu theorem (2. 4) defines every continuous circulation function on σ in terms of its boundary values and its areal derivative on σ. The question then is: how must $F(z)$ be chosen so that $f(u)$ defined by the equation

$$(5.1) \qquad \frac{1}{2\pi i}\int_c \frac{F(z)\,dz}{z-u} - \frac{1}{\pi}\int_\sigma \frac{\frac{\partial F}{\partial \alpha}\,d\sigma}{z-u} \equiv f(u)$$

is a continuous circulation function on σ? The answer to this question lies in the following

THEOREM 2. *The necessary and sufficient condition for the existence of a continuous circulation function whose areal derivative $\partial F/\partial\alpha$ is given on σ and which assumes the values F on c is that F satisfies the functional equation* (4. 3).

That the condition is necessary is proved in § 4 and stated in Theorem 1.

To show that the condition is sufficient one may first show that the function $f(u)$ defined by (5. 1) assumes the value $F(x)$ on the boundary under the hypothesis that $F(x)$ satisfies equation (4. 3), that is

$$(5.2) \qquad \frac{1}{2\pi i}\int_c \frac{F(z)\,dz}{z-x} - \frac{1}{\pi}\int_\sigma \frac{\frac{\partial F}{\partial \alpha}\,d\sigma}{z-x} = \tfrac{1}{2}F(x).$$

When u is made to approach x of c (5. 1) becomes

$$\phi_i(F) - \frac{1}{\pi}\int_\sigma \frac{\frac{\partial F}{\partial \alpha}\,d\sigma}{z-x} = f(x)$$

in the limit. If $\phi_i(F)$ which exists is replaced by its value from the first of (3. 1) and the area integral by its value from the hypothesis (5. 2) one obtains the identity

$$F(x) \equiv f(x).$$

That $f(u)$ of (5.1) is a circulation function in σ may be verified by multiplying (5.1) by du and integrating around a closed contour c' inside c bounding σ'. The first integral in (5.1), a function of u, has z on the boundary c; thus $1/(z-u)$ has no poles in σ. When the order of integration is interchanged the integral of $1/(z-u)$ around c' therefore vanishes. Hence

$$-\frac{1}{\pi}\int_{\sigma}\frac{\partial F}{\partial \bar{z}}\,d\sigma\int_{c'}\frac{du}{z-u}=\int_{c'}f(u)\,du;$$

and since now z, a point of σ, is a simple pole,

$$\int_{c'}\frac{du}{z-u}=-2\pi i$$

when z is a point of σ' and zero when outside c'; it follows that

$$\frac{1}{\pi}\int_{\sigma'}\frac{\partial F}{\partial \bar{z}}\,d\sigma=\frac{1}{2\pi i}\int_{c'}f(u)\,du.$$

When one divides this equation by σ' and takes the limit as c' is contracted to a point one finds in the usual way that

$$\partial F/\partial \bar{z}=\partial f/\partial \bar{z}.$$

A resubstitution of $\partial f/\partial \bar{z}$ for $\partial F/\partial \bar{z}$ in the previous equation defines $f(u)$ as a circulation function.

As a consequence of Theorem 2 one has the

COROLLARY. *The necessary and sufficient condition for the existence of monogenic function F on σ is that F satisfies the functional equation*

$$\frac{1}{2\pi i}\int_{c}\frac{f(z)\,dz}{z-x}=\tfrac{1}{2}f(x).$$

This corollary, a theorem by Picard,[6] follows directly from the above theorem, since the vanishing of the areal derivative defines the function as monogenic on σ.

However, the deduction of the theorem is invalid if c is any part of the boundary of the total region D in which the circulation function is defined since in the deduction the Pompeiu theorem is applied to the area between o_2 and c outside c. To treat this case assume that a continuous circulation function exists in σ *whose areal derivative is $\partial F/\partial \bar{z}$* on σ and which assumes the value F on c; c bounding σ is now all or a part of the boundary of D. Then for an area σ' bounded by a curve c', a neighboring curve to c in σ, (2.4) is valid or

---

[6] E. Picard, *Quelques types d'équations aux dérivées partielles.* Gauthier-Villars (1927), p. 67.

$$\frac{1}{2\pi i}\int_{c'}\frac{f(z)\,dz}{z-u}=\frac{1}{\pi}\int_{\sigma'}\frac{\frac{\partial f}{\partial \bar{z}}\,d\sigma}{z-u}+f(u)$$

where u is a point of σ'. If c' is made to approach c, then since $\partial f/\partial \bar{z}=\partial F/\partial \bar{z}$ on σ and $f=F$ on c one has in the limit

$$(5.2)\qquad \frac{1}{2\pi i}\int\frac{F(z)\,dz}{z-u}=\frac{1}{\pi}\int_{\sigma}\frac{\frac{\partial F}{\partial \bar{z}}\,d\sigma}{z-u}+f(u).$$

Now let u approach x of c; this last equation in the limit becomes

$$\phi_i(F)=\frac{1}{\pi}\int_{\sigma}\frac{\frac{\partial F}{\partial \bar{z}}\,d\sigma}{z-x}+F(x).$$

Had u been taken outside c, (2. 3) would have been applicable so that

$$\frac{1}{2\pi i}\int_{c}\frac{F(z)\,dz}{z-u}=\frac{1}{\pi}\int_{\sigma}\frac{\frac{\partial F}{\partial \bar{z}}\,d\sigma}{z-u}$$

and in the limit as u approaches x one finds that

$$\phi_e(F)=\frac{1}{\pi}\int_{\sigma}\frac{\frac{\partial F}{\partial \bar{z}}\,d\sigma}{z-x}.$$

Therefore ϕ_i and ϕ_e exist and by the Priwaloff Theorem [7] so does the principal value.

When $\phi_i(F)$ is replaced by its value from the first of (3. 1) one gets

$$\frac{1}{2\pi i}\int\frac{F(z)\,dz}{z-x}=\frac{1}{\pi}\int_{\sigma}\frac{\frac{\partial F}{\partial \bar{z}}\,d\sigma}{z-x}+\tfrac{1}{2}F(x),$$

which again proves the necessity of the condition. The sufficiency is then proved as in the previous case.

6. The exterior problem. Circulation functions in an infinite region D outside a bounding curve will be studied in this section. Let c be the boundary of an infinite area σ in D. It will now be possible to extend the Pompeiu theorem.

Consider the annular area σ'' between c and a large circle O radius R outside c. Let $f(z)$ be a continuous circulation function in the region D which takes the value p at infinity.

[7] Priwaloff, *loc. cit.*

If u is a point inside the curve c, that is, outside the infinite area, then (2. 3) generalizes for the annular area σ'' into

$$(6.1)\qquad \frac{1}{2\pi i}\int_{o}\frac{f(z)\,dz}{z-u}-\frac{1}{2\pi i}\int\frac{f(z)\,dz}{z-u}=\frac{1}{\pi}\int_{\sigma''}\frac{\frac{\partial f}{\partial \alpha}\,d\sigma}{z-u}.$$

The integral

$$\int_{o}\frac{f(z)\,dz}{z-u}=p\int_{o}\frac{dz}{z-u}+\int_{o}\frac{f(z)-p}{z-u}\,dz.$$

The first integral on the right here is $2\pi ip$ while the second vanishes with increasing R, so that

$$(6.2)\qquad p-\frac{1}{2\pi i}\int_{c}\frac{f(z)\,dz}{z-u}=\frac{1}{\pi}\int_{\sigma}\frac{\frac{\partial f}{\partial \alpha}\,d\sigma}{z-u}.$$

If u is a point of the annular area and the center of a small circle o radius r, (6. 1) becomes

$$\frac{1}{2\pi i}\int_{O}\frac{f(z)\,dz}{z-u}-\frac{1}{2\pi i}\int_{c}\frac{f(z)\,dz}{z-u}-\frac{1}{2\pi i}\int_{O}\frac{f(z)\,dz}{z-u}=\frac{1}{\pi}\int_{\sigma'}\frac{\frac{\partial f}{\partial \alpha}\,d\sigma}{z-u}$$

where σ' is that portion of the annular area outside o. In extending this result to the infinite area note that the first integral on the left has been evaluated while

$$\int_{o}\frac{f(z)\,dz}{z-u}=f(u)\int_{o}\frac{dz}{z-u}+\int_{o}\frac{f(z)-f(u)}{z-u}\,dz.$$

The first integral on the right is evidently $2\pi i\,f(u)$ while the second vanishes with r because of the continuity of $f(z)$. For

$$\left|\int_{o}\frac{f(z)-f(u)}{z-u}\,dz\right|\leqq\int_{o}|f(z)-f(u)|\,d\theta$$

while if M and m are the maximum and minimum values of $|f(z)-f(u)|$ respectively, it follows that

$$2\pi m\leqq\int_{o}|f(z)-f(u)|\,d\theta\leqq 2\pi M.$$

Since $m\rightarrow M\rightarrow 0$ as $r\rightarrow 0$, the truth of the statement will be evident. Thus for the infinite area σ one finds that

$$(6.3)\qquad p-\frac{1}{2\pi i}\int_{c}\frac{f(z)\,dz}{z-u}-f(u)=\frac{1}{\pi}\int_{\sigma}\frac{\frac{\partial f}{\partial \alpha}\,d\sigma}{z-u}.$$

This is the extension of the Pompeiu Theorem. Therefore $f(u)$, a continuous circulation function on the infinite area σ, may be expressed in terms of its areal derivative over the area, its value at infinity and its boundary values in the form

$$(6.4) \qquad f(u) = p - \frac{1}{2\pi i}\int_c \frac{f(z)\,dz}{z-u} - \frac{1}{\pi}\int_\sigma \frac{\frac{\partial f}{\partial \alpha}\,d\sigma}{z-u}.$$

If the boundary c of the infinite area σ is entirely within D and if c has a unique tangent almost everywhere, then the existence of the principal value of the singular integral for a continuous circulation function on σ can be deduced by the use of (6.2) and the extended Pompeiu theorem (6.4).

The question to be considered here is the following: Under what conditions does $f(z)$ given by the equation

$$(6.5) \qquad p - \frac{1}{2\pi i}\int \frac{F(z)\,dz}{z-u} - \frac{1}{\pi}\int_\sigma \frac{\frac{\partial f}{\partial \alpha}\,d\sigma}{z-u} \equiv f(u)$$

define a continuous circulation function on σ?

If c is a contour entirely in the infinite region where $F(z)$ is a continuous circulation function, then $f(z) \equiv F(z)$. Under this condition let x on c be the center of a small circle radius r. If the notation of § 4 be used here, then, for the infinite area bounded by $c + o_2$, (6.4) is applicable so that

$$(6.6) \qquad p - \frac{1}{2\pi i}\int\limits_{c_1+o_2} \frac{F(z)\,dz}{z-x} - \frac{1}{\pi}\int\limits_{\sigma+\sigma_2} \frac{\frac{\partial F}{\partial \alpha}\,d\sigma}{z-x} = F(x).$$

For the infinite area bounded by $c_1 + o_1$ one has

$$(6.7) \qquad p - \frac{1}{2\pi i}\int\limits_{c_1+o_1} \frac{F(z)\,dz}{z-x} - \frac{1}{\pi}\int\limits_{\sigma-\sigma_1} \frac{\frac{\partial F}{\partial \alpha}\,d\sigma}{z-x} = 0.$$

If (6.6) and (6.7) are added, the integrals along o_1 and o_2 cancel each other as in the finite case, while $c_1 \to c$ and $\sigma + \sigma_2 \to \sigma - \sigma_1 \to \sigma$ so that in the limit

$$(6.8) \qquad p - \frac{1}{2\pi i}\int_c \frac{F(z)\,dz}{z-x} - \frac{1}{\pi}\int_\sigma \frac{\frac{\partial F}{\partial \alpha}\,d\sigma}{z-x} = \tfrac{1}{2}F(x).$$

Thus if $F(z)$ is a continuous circulation function in D, it necessarily follows that $F(z)$ satisfies the functional equation (6.8). The sufficiency of

the condition may be proved as in the finite case. Incidentally, the existence of the principal value has been demonstrated.

The second case treated in § 5 arises here. Omitting the details one may summarize the results in

THEOREM 3. *The necessary and sufficient condition for the existence of a continuous circulation function F on the infinite area σ, whose areal derivative is $\partial F/\partial\alpha$ on σ, whose value at infinity is p and which assumes the value $F(x)$ on the boundary c, is that F satisfies the functional equation* (6.5), and the

COROLLARY. *The necessary and sufficient condition for the existence of monogenic function F on the infinite area σ, whose value at infinity is p and which assumes the values F on the boundary c is that F satisfies the functional equation*

$$\tfrac{1}{2}F(x) = p - \frac{1}{2\pi i}\int_c \frac{F(z)dz}{z-x}.$$

7. Circulation functions of the second kind. Returning to the finite area σ bounded by c, an anomalous situation arises for the integral $\int_c \frac{f(z)d\bar{z}}{\bar{z}-\bar{u}}$ in that u is a simple pole of the integrand though it does not appear explicitly as such. For when $z = u$, $\bar{z} = \bar{u}$. If the Kasner mean derivative[8] $\partial f/\partial\beta$ is introduced, *circulation functions of the second kind* arise. They are defined as polygenic functions $f(z)$ for which the circulation theorem

$$\frac{1}{2\pi i}\int_c f(z)d\bar{z} = -\frac{1}{\pi}\int_\sigma \frac{\partial f}{\partial\beta}d\sigma \tag{7.1}$$

is valid for all rectifiable closed curves in a region D.

If u is outside c, (7.1) generalizes into

$$\frac{1}{2\pi i}\int_c \frac{f(z)d\bar{z}}{\bar{z}-\bar{u}} = -\frac{1}{\pi}\int_\sigma \frac{\frac{\partial f}{\partial\beta}d\sigma}{\bar{z}-\bar{u}} \tag{7.2}$$

while the analogue to the Pompeiu theorem,

$$\frac{1}{2\pi i}\int_c \frac{f(z)d\bar{z}}{\bar{z}-\bar{u}} = -\frac{1}{\pi}\int_\sigma \frac{\frac{\partial f}{\partial\beta}d\sigma}{\bar{z}-\bar{u}} - f(u) \tag{7.3}$$

is easily deduced.

[8] Kasner, "General theory of polygenic or non-monogenic functions. The derivative congruence of circles," *Proceedings of the National Academy of Sciences* (January, 1928), pp. 75-82.

The integral

$$\frac{1}{2\pi i}\int_c \frac{f(z)d\bar{z}}{\bar{z}-\bar{u}}$$

becomes singular when u becomes a point x on c. If

$$\frac{1}{2\pi i}\int_c \frac{f(z)d\bar{z}}{\bar{z}-\bar{x}}$$

is written for the principal value of this integral and ψ_i and ψ_c for the other limit values respectively as indicated by the subscripts it may be shown that if the principal value of this singular integral exists the other limit values exist and conversely. Omitting details, one may express the relations of these limit values by the equations

$$(7.4)\qquad \begin{aligned} \psi_i &= \frac{1}{2\pi i}\int_c \frac{f(z)d\bar{z}}{\bar{z}-\bar{x}} - \tfrac{1}{2}f(x) \\ \psi_c &= \frac{1}{2\pi i}\int_c \frac{f(z)d\bar{z}}{\bar{z}-\bar{x}} + \tfrac{1}{2}f(x). \end{aligned}$$

By the use of equations (7.2), (7.3), and (7.4) a complete analogue to the results obtained in the previous sections may be found for circulation functions of the second kind, the analogue to the monogenic function being the antimonogenic function, a function monogenic in $\bar{z}$. Extensions to the infinite area may then be made.

UNIVERSITY OF MICHIGAN.

FURTHER GENERALIZATIONS OF THE CAUCHY INTEGRAL FORMULA.*[1]

By Dawson G. Fulton.[2]

1. The generalizations given in this paper arise out of a consideration of the structure of the Cauchy Integral Formula when it is freed of the special form in which it appears due to the use of complex numbers. Thus in the formula

$$f(a)=\frac{1}{2\pi i}\int_C \frac{f(z)\,dz}{z-a}, \tag{1}$$

we set $a=0$, $f(z)=Y+iX$, $z=x+iy$, $dz=dx+idy$, $-i(dx+idy)=(\alpha+i\beta)ds$, where α and β are the direction cosines of the normal to the contour; then by equating real and imaginary parts we get

$$\begin{aligned} 2\pi X_0 &= \int_C \left\{ X\frac{x\alpha+y\beta}{r^2}+Y\frac{x\beta-y\alpha}{r^2}\right\} ds \\ 2\pi Y_0 &= \int_C \left\{ X\frac{y\alpha-x\beta}{r^2}+Y\frac{x\alpha+y\beta}{r^2}\right\} ds, \qquad (r^2=x^2+y^2). \end{aligned} \tag{2}$$

We notice that the integrand is linear and homogeneous in the components X and Y, and also in the direction cosines α and β, and homogeneous of degree -1 in the independent variables x and y. More specifically we say that the integrand is linear and homogeneous in the independent variables besides containing in the denominator a quadratic form of the independent variables.

We choose these observations as the starting point of our generalizations, and propose the following considerations. Is it possible, and, if so, under what conditions, to find a general integral formula of this same structure which can be used for the same purpose as is the Cauchy Integral Formula; that is, for the calculations of the values of functions in the interior points of a region when their values are given on the boundary? Using the index notation, denoting X by X^1, Y by X^2, x by x^1, y by x^2, etc., summations by Greek letters used as subscripts and superscripts, we write the integral we wish to consider as

* Received November 28, 1938.

[1] Presented before the American Mathematical Society, Chicago, April, 1937.

[2] The writer wishes to express his appreciation to G. Y. Rainich, of the University of Michigan, for his helpful suggestions during the preparation of this paper.

$$\int \frac{1}{r^2} N^{i\nu}_{\sigma\lambda} x^\sigma X^\lambda \alpha_\nu ds, \tag{3}$$

where the constants N^{ij}_{kl} are arbitrarily chosen, and $r^2 = g_{\alpha\beta}x^\alpha x^\beta$ is a positive definite form.

The original formula (2) is obtained from this by setting

$$N^{ij}_{kl} = \delta_{il}\delta_{jk} + \delta_{ik}\delta_{jl} - \delta_{ij}\delta_{kl}; \qquad g_{ij} = \delta_{ij}, \tag{4}$$

where δ_{ij} etc. mean the Kronecker delta. δ_{ij} wherever used throughout the paper will have the same meaning.

The formulas (2) are applicable to analytic functions; hence in our consideration to a pair of functions X and Y, which are differentiable (possess Stolz differentials) and satisfy the conditions—the Cauchy-Riemann differential equations—

$$\frac{\partial X^1}{\partial x^1} + \frac{\partial X^2}{\partial x^2} = 0, \qquad \frac{\partial X^1}{\partial x^2} - \frac{\partial X^2}{\partial x^1} = 0. \tag{5}$$

The proof that the formulas (2) hold for functions which satisfy (5) is based on the following points: (1) that the integrals are independent of the particular contour about which they are taken as long as they surround the point at which we wish to make the evaluation, (2) that calculations are made of integrals over a circle around the point in question, and (3) that the desired evaluation is obtained by passing to the limit by making the radius of the circle go to zero. We want to know what conditions must be imposed on the general formula (3) so that the same results can be obtained.

2. In discussing these conditions let us consider the special case where $g_{ij} = \delta_{ij}$. First of all we rewrite formula (3) in a form which will evaluate the functions X and Y at any inner point a : (a^1, a^2), rather than at the origin. Formula (3) then becomes

$$(X^i)_a = \frac{1}{2\pi}\int_C \frac{1}{r^2} N^{i\nu}_{\sigma\lambda} (x^\sigma - a^\sigma) X^\lambda \alpha_\nu ds, \tag{6}$$

where by $(X)_a$ we mean X evaluated at the point a, and

$$r^2 = (x^\alpha - a^\alpha)(x^\alpha - a^\alpha).$$

Since the integral is to be independent of the contour about which it is taken, we know that

$$\frac{\partial}{\partial x^\nu} \text{(coefficient of } \alpha_\nu) = 0; \tag{7}$$

that is,

$$(x^\alpha - a^\alpha)(x^\alpha - a^\alpha)(x^\sigma - a^\sigma)N_{\sigma\lambda}^{i\nu}\frac{\partial X^\lambda}{\partial x^\nu} + (x^\alpha - a^\alpha)(x^\alpha - a^\alpha)N_{\sigma\lambda}^{i\sigma}X^\lambda$$
$$- 2(x^\sigma - a^\sigma)(x^\nu - a^\nu)N_{\sigma\lambda}^{i\nu}X^\lambda = 0.$$

This equation must be true for all points a; hence it will hold for ta where t takes on all values such that ta is a point within R. This will give a cubic equation in t which must be true identically; that is, the coefficient of each term must vanish. Writing down the coefficient of t^3, we get

$$a^\alpha a^\alpha a^\sigma N_{\sigma\lambda}^{i\nu}\frac{\partial X^\lambda}{\partial x^\nu} = 0.$$

This equation must be true identically in a, thus

$$N_{k\lambda}^{i\nu}\frac{\partial X^\lambda}{\partial x^\nu} = 0. \tag{8}$$

Applying this formula (8) to the term which is free of t; namely,

$$x^\alpha x^\alpha x^\sigma N_{\sigma\lambda}^{i\nu}\frac{\partial X^\lambda}{\partial x^\nu} + x^\alpha x^\alpha N_{\sigma\lambda}^{i\sigma}X^\lambda - 2x^\sigma x^\nu N_{\sigma\lambda}^{i\nu}X^\lambda = 0,$$

we get

$$x^\alpha x^\alpha N_{\sigma\lambda}^{i\sigma}X^\lambda - 2x^\sigma x^\nu N_{\sigma\lambda}^{i\nu}X^\lambda = 0.$$

This equation holds for all X^i and x^i; hence we get

$$2\delta_{jk}N_{\sigma l}^{i\sigma} = 2(N_{kl}^{ij} + N_{jl}^{ik}). \tag{9}$$

By an easy substitution we find that for $j = k = 1$ or 2, $N_{1l}^{i1} = N_{2l}^{i2}$, and that for $j \neq k$, $N_{kl}^{ij} = -N_{jl}^{ik}$. The conditions (8) and (9) assure the vanishing of the coefficients of t^2 and t. Thus we see that (8) and (9) are sufficient conditions that the integral formula (6) be independent of the contour.

Let us assume the conditions (8) and (9) and then we may integrate the formulas (6) about any contour whatsoever. We shall choose a circle with center at the given point a. Then

$$(x^1 - a^1)/r = \alpha, \qquad (x^2 - a^2)/r = \beta. \tag{10}$$

Using (9) and (10) the formulas (6) become

$$(X^i)_a = \frac{1}{2\pi}\int_C (N_{11}^{i1}X^1 + N_{12}^{i1}X^2)\frac{ds}{r^2}.$$

But this integral is independent of r, and we may write

$$\frac{1}{2\pi}\int_C (N^{i1}_{11}X^1 + N^{i1}_{12}X^2)\frac{ds}{r^2} = N^{i1}_{11}(X^1)_a + N^{i1}_{12}(X^2)_a.$$

Since this must give the required evaluation $(X^i)_a$, we see that when $i=1$

$$(X^1)_a = N^{11}_{11}(X^1)_a + N^{11}_{12}(X^2)_a.$$

Hence $N^{11}_{11}=1$ and $N^{11}_{12}=0$. Similarly when $i=2$, $N^{21}_{11}=0$, $N^{21}_{12}=1$. From these values, together with formulas (9), it is easily verified that

$$N^{i\sigma}_{\sigma l} = 2\delta_{il}, \tag{11}$$

and therefore

$$N^{ij}_{kl} + N^{ik}_{jl} = \delta_{jk}N^{i\sigma}_{\sigma l} = 2\delta_{jk}\delta_{il}. \tag{12}$$

Thus we see that (8) and (12) are sufficient conditions which must be imposed upon the functions X and Y and the constants N^{ij}_{kl} in order to insure the fact that the formulas (6) will evaluate the given functions at any interior point of a region in terms of the boundary values of the functions.

3. One other point must be settled, however, before we state the theorem. The system of differential equations (8) consists of four linear homogeneous equations with constant coefficients. In order for this system to have a solution other than the trivial case where the functions are all constants, the determinant of the matrix N^{ij}_{kl} must be equal to zero. This means, of course, that there must be some relationship among the equations (8), and suggests that it might be possible to find an equivalent set of fewer equations. The number of this latter set will, of course, depend upon the rank of the matrix N^{ij}_{kl}. If one writes out the matrix using (11) and (12) it is obvious that the rank cannot be one unless we use pure imaginaries for the constants N^{ij}_{kl}, and we do not wish to do this. Since the equations do not reduce to one equation, then, let us look at the next most interesting case; namely, when the rank of the matrix is 2. If we write out the first two equations for which $i=1$, we find

$$\frac{\partial X}{\partial x} + \frac{\partial(qX+sY)}{\partial y} = 0, \qquad \frac{\partial X}{\partial y} - \frac{\partial(qX+sY)}{\partial x} = 0, \tag{13}$$

where $N^{12}_{11} = -N^{11}_{21} = q$, $N^{12}_{12} = -N^{11}_{22} = s$. These are the Cauchy-Riemann differential equations for the functions X and $qX+sY$. Because of these equations then we may write integral formulas for X and $qX+sY$ exactly as we did for X and Y in (2). Thus we find $(X)_a$ and $(Y)_a$ as a result of the first two equations only. That is, while the four equations in (8) are sufficient, the two for which $i=1$ or 2 are all which are necessary for

the possibility of the desired evaluations. We note also that for this set-up either of the two pairs of equations will also insure the independence conditions which we discussed in section 2. Thus we see that the set of equations (8) breaks up into two sets of two equations (those obtained by setting $i=1$, or those for which $i=2$), either one of which is sufficient for our purposes. If we have one of these pairs of equations given and we wish to evaluate both $(X)_a$ and $(Y)_a$ by using formula (6) only, then we shall have to make the two pairs of equations equivalent. This means, as is readily checked, that we impose the further condition on the constants N^{ij}_{kl}

$$(14) \qquad N^{ij}_{ii} = N^{ji}_{jj}, \quad (i \neq j); \qquad \frac{1+(N^{ii}_{ij})^2}{N^{ij}_{ij}} = N^{ji}_{ji}, \quad (i \neq j).$$

This is quite similar to the reduction of the equations (8) when we give to N^{ij}_{kl} the evaluation (4). Using this evaluation we get two sets of two equations each, and the sets are identical. In the case which we have discussed in this section one set may be transformed into the other by linear transformations on the functions only.

We now state the following *theorem:*

Given any region R bounded by C, the functions X and Y may be evaluated at any inner point a of R in terms of their values on C by means of the integral formulas

$$(X^i)_a = \frac{1}{2\pi}\int_C \frac{1}{r^2} N^{i\nu}_{\sigma\lambda}(x^\sigma - a^\sigma)X^\lambda \alpha_\nu ds; \quad r^2 = (x^a - a^a)(x^a - a^a),$$

provided the functions X and Y possess a Stolz differential and satisfy within the region R the system of differential equations

$$N^{i\nu}_{k\lambda}\frac{\partial X^\lambda}{\partial x^\nu} = 0,$$

and the constants N^{ij}_{kl} satisfy the conditions

$$N^{ij}_{kl} + N^{ik}_{jl} = 2\delta_{jk}\delta_{il},$$

and for $i \neq j$

$$N^{ij}_{ii} = N^{ji}_{jj}, \quad \frac{1+(N^{ii}_{ij})^2}{N^{ij}_{ij}} = N^{ji}_{ji}.$$

4. In sections 2 and 3 we have demonstrated the conditions of which we inquired in section 1, but so far we have considered only the special case where

$$r^2 = (x^a - a^a)(x^a - a^a).$$

We wish now to discuss the more general expression for (6) when we use for r^2 the positive definite form

$$g_{\alpha\beta}(x^\alpha - a^\alpha)(x^\beta - a^\beta).$$

We wish to find the conditions which we must impose in this case. If we carry through the same procedure here as we did in section 2, we find the conditions for independence of contour to be that the functions must satisfy

$$(15) \qquad N^{i\nu}_{k\lambda}\frac{\partial X^\lambda}{\partial x^\nu} = 0,$$

where

$$g_{\nu j}N^{i\nu}_{kl} + g_{\nu k}N^{i\nu}_{jl} = g_{jk}N^{i\sigma}_{\sigma l}.$$

These and subsequent conditions could be obtained immediately from the results of sections 2 and 3 if we could find a linear transformation on the variables which would carry the form $(x^\alpha - a^\alpha)(x^\alpha - a^\alpha)$ into $g_{\alpha\beta}(x^\alpha - a^\alpha)(x^\beta - a^\beta)$ for any given set of g_{ij}'s. Such a transformation does exist, we know, and is found by choosing as the constants c_{ij} in the transformation the components of two vectors the square of whose lengths are the positive numbers g_{11}, and g_{22}, g_{12} being the scalar product of the two; that is, $g_{ij} = c_{\lambda i}c_{\lambda j}$. Let us apply this transformation, $x^i = c'_{\rho i}(x^\rho)'$, $(x^i)' = c_{i\rho}x^\rho$, to the variables in the formulas (8). We get

$$(17) \qquad \int \frac{1}{r^2} c_{\rho\sigma}c'_{\tau\nu}N^{i\tau}_{\rho\lambda}(x^\sigma - a^\sigma)X^\lambda\alpha_\nu ds,$$

which we write as

$$\int \frac{1}{r^2} M^{i\nu}_{\sigma\lambda}(x^\sigma - a^\sigma)X^\lambda\alpha_\nu ds,$$

where

$$M^{ij}_{kl} = c_{\rho k}c'_{\tau j}N^{i\tau}_{\rho l}; \quad r^2 = g_{\alpha\beta}(x^\alpha - a^\alpha)(x^\beta - a^\beta); \quad g_{ij} = c_{\lambda i}c_{\lambda j}.$$

But by (16) we know that

$$g_{jk}M^{i\tau}_{\sigma l} = g_{\nu j}M^{i\nu}_{kl} + g_{\nu k}M^{i\nu}_{jl}.$$

That is,

$$c_{\lambda j}c_{\lambda k}M^{i\sigma}_{\sigma l} = c_{\lambda\nu}c_{\lambda j}M^{i\nu}_{kl} + c_{\lambda\nu}c_{\lambda k}M^{i\nu}_{jl}.$$

Therefore

$$c_{\lambda\rho}c'_{j\rho}c_{\lambda\tau}c'_{k\tau}M^{i\sigma}_{\sigma l} = c'_{j\rho}c'_{k\tau}c_{\lambda\nu}c_{\lambda\rho}M^{i\nu}_{\tau l} + c'_{j\rho}c'_{k\tau}c_{\lambda\nu}c_{\lambda\tau}M^{i\nu}_{\rho l},$$

or

$$\delta_{jk}M^{i\sigma}_{\sigma l} = c'_{k\tau}c_{j\nu}M^{i\nu}_{\tau l} + c'_{j\rho}c_{k\nu}M^{i\nu}_{\rho l}, \text{ since } c'_{j\rho}c_{k\rho} = \delta_{jk}.$$

But

$$M^{i\sigma}_{\sigma l} = N^{i\sigma}_{\sigma l}, \text{ and } c'_{j\tau}c_{k\nu}M^{i\nu}_{\tau l} = N^{ij}_{kl}.$$

Hence we get

$$\delta_{jk}N^{i\sigma}_{\sigma l} = N^{ik}_{jl} + N^{ij}_{kl}.$$

But this is exactly equation (9) found in section 2. Thus we have shown that this general condition is derivable from the special case of section 2 and 3 by means of linear transformations on the variables. It is therefore possible to apply the reasoning of those sections to this general case. We get then

$$g_{\nu j} N^{i\nu}_{kl} + g_{\nu k} N^{i\nu}_{jl} = g_{jk} N^{i\sigma}_{\sigma l} = 2 g_{jk} \delta_{il} \tag{18}$$

as the conditions to be satisfied by the constants N^{ij}_{kl} and g_{ij} for this case.

This brings us then to the following theorem:

Given any region R bounded by C, the functions X and Y may be evaluated at any inner point a of R in terms of their values on C by means of the integral formulas

$$\int \frac{1}{r^2} N^{i\nu}_{\sigma\lambda}(x^\sigma - a^\sigma) X^\lambda \alpha_\nu ds; \quad r^2 = g_{\alpha\beta}(x^\alpha - a^\alpha)(x^\beta - a^\beta),$$

provided the functions X and Y possess a Stolz differential and satisfy within the region R the system of differential equations

$$N^{i\nu}_{k\lambda} \frac{\partial X^\lambda}{\partial x^\nu} = 0,$$

and the constants N and g satisfy the conditions

$$g_{\nu j} N^{i\nu}_{kl} + g_{\nu k} N^{i\nu}_{jl} = 2 g_{jk} \delta_{il}.$$

Here as before the determinant of the matrix N^{ij}_{kl} must be zero. The additional conditions necessary to have an equivalent set of two equations are not developed here.

To verify the fact that this does give the desired values, one integrates about the ellipse $g_{\alpha\beta}(x^\alpha - a^\alpha)(x^\beta - a^\beta) = K$.

5. Instead of starting with an integral representation, it is possible to approach this problem by considering two differential equations of the form

$$\begin{aligned} A\frac{\partial X}{\partial x} + B\frac{\partial X}{\partial y} + C\frac{\partial Y}{\partial x} + D\frac{\partial Y}{\partial y} &= 0 \\ A'\frac{\partial X}{\partial x} + B'\frac{\partial X}{\partial y} + C'\frac{\partial Y}{\partial x} + D'\frac{\partial Y}{\partial y} &= 0. \end{aligned} \tag{19}$$

This system was treated by H. C. Chang in his thesis on "Transformations of Linear Partial Differential Equations."[3] He showed that the general function $X + iY$ could be reduced to an analytic function when X and Y

[3] See H. C. Chang, "Transformation of partial differential equations," *Science Society of China*, vol. 7 (1931), no. 2.

satisfied the system of differential equations (19), provided the functions X and Y are differentiable, and that the following condition on the coefficients of (19) is satisfied

$$(20) \quad 4(BD'-B'D)(AC'-A'C)-(AD'-A'D+BC'-B'C)^2>0.$$

We ask now whether it is possible to find an integral formula analogous to the Cauchy Integral formula whereby we shall be able to evaluate this general function at some inner point of a region in terms of its values on the boundary. Such a formula was obtained by applying the same linear transformations to the functions and variables in the Cauchy Integral Formula as were applied to the Cauchy-Riemann Differential Equations to secure the system (19). The formula has as its constants certain expressions involving the coefficients of (19). These expressions are combinations of M_{ij}'s, where by M_{ij} we mean the determinant formed by the i-th and the j-th columns of the matrix

$$(21) \qquad \begin{Vmatrix} A & B & C & D \\ A' & B' & C' & D' \end{Vmatrix}.$$

As a convenience in writing, the following notations have been introduced

$$(22.1) \qquad R^2=4M_{13}M_{24}-(M_{14}+M_{23})^2,$$

$$(22.2) \qquad g=M_{24}x^2-(M_{14}+M_{23})xy+M_{13}y^2.$$

Using this notation we now state the following *theorem:*

If R is a given region, of which C is the boundary, then the functions X and Y may be evaluated at the origin by means of the formulas

$$(23.1) \quad 2\pi X_0=\int_C \frac{1}{2Rg}[\{(M_{14}-M_{23})X-2M_{34}Y\}dg+R^2(x\alpha+y\beta)Xds],$$

$$(23.2) \quad 2\pi Y_0=-\int_C \frac{1}{2Rg}[\{2M_{12}X+(M_{14}-M_{23})Y\}dg-R^2(x\alpha+y\beta)Yds],$$

provided the functions X and Y are differentiable and satisfy the differential equations

$$A\frac{\partial X}{\partial x}+B\frac{\partial X}{\partial y}+C\frac{\partial Y}{\partial x}+D\frac{\partial Y}{\partial y}=0$$

$$A'\frac{\partial X}{\partial x}+B'\frac{\partial X}{\partial y}+C'\frac{\partial Y}{\partial x}+D'\frac{\partial Y}{\partial y}=0,$$

whose coefficients satisfy the inequality (20).

Obviously a translation of the axes will make this formula applicable to any inner point of the region considered.

The integral formulas (23), together with the differential equations (19), are derived in the same manner as were the corresponding formulas and equations in section 4; namely, by means of linear transformations. In this case we have two linear differential equations instead of four. This, then, is clearly a special case of that considered in section 4.

6. The discussions of sections 2, 3 and 4 dealt entirely with the case of two dimensions. Integral formulas which are generalizations of the Cauchy Integral Formula for analytic functions of higher dimensions have been obtained, and the results appear in a joint paper by G. Y. Rainich and the writer.[4]

We wish to consider now the extension within these higher dimensions similar to the one which we have given for two dimensions. For the general case of n dimensions we study the integral formulas

$$\frac{2\pi^{\frac{1}{2}n}}{\Gamma(\frac{1}{2}n)}(X^i)_a = \int \frac{1}{r^n} N^{i\nu}_{\sigma\lambda}(x^\sigma - a^\sigma)X^\lambda \alpha_\nu dw, \tag{24}$$

where $r^2 = g_{\alpha\beta}(x^\alpha - a^\alpha)(x^\beta - a^\beta)$, and all summations are from 1 to n. We want to know what conditions must be imposed here in order that we may evaluate the n functions at any inner point of an n-dimensional region in terms of their values on a hyper-surface which bounds the region. The arguments of the previous sections carry through here very readily, and we may state the following general *theorem:*

If T is a given n-dimensional region bounded by a hypersurface S, then we may evaluate the functions X^i $(i = 1, 2, \cdots n)$ at any inner point $a: (a^1, a^2, \cdots, a^n)$ in terms of their values on the hypersurface by means of the integral formulas

$$\frac{2\pi^{\frac{1}{2}n}}{\Gamma(\frac{1}{2}n)}(X^i)_a = \int \frac{1}{r^n} N^{i\nu}_{\sigma\lambda}(x^\sigma - a^\sigma)X^\lambda \alpha_\nu dw, \tag{25}$$

where $r^2 = g_{\alpha\beta}(x^\alpha - a^\alpha)(x^\beta - a^\beta)$, provided the functions X^i possess Stolz differentials and satisfy within the region T the following system of differentiable equations

$$N^{i\nu}_{k\lambda}\frac{\partial X^\lambda}{\partial x^\nu} = 0, \tag{26}$$

together with the conditions that

$$n(g_{\nu j}N^{i\nu}_{kl} + g_{\nu k}N^{i\nu}_{jl}) = 2g_{jk}N^{i\sigma}_{\sigma l} = 2ng_{jk}\delta_{il}. \tag{27}$$

---

[4] G. Y. Rainich and D. G. Fulton, "Generalizations of the Cauchy integral formula to higher dimensions," *American Journal of Mathematics*, vol. 54 (1932), pp. 235-242.

The formulas derived earlier in the paper are clearly special cases of this general set-up. By setting $n = 2$, we get the formulas and corresponding conditions of section 4, and for $n = 2$ and $g_{ij} = \delta_{ij}$ we get the results of sections 2 and 3.

It might further be pointed out that the formulas given in the joint paper referred to above are also special cases of this general formula (25); they may be obtained by giving to n the proper value, and to N^{ij}_{kl} and g_{ij} the evaluation given in (4) section 1. Likewise formula (26), under the same substitution for N^{ij}_{kl}, yields the corresponding set of differential equations for each case. It will be noted that for this special evaluation of the constants N^{ij}_{kl} the number of differential equations (26) reduces to $(n^2 - n + 2)/2$.

It should be pointed out that the reduction of the number of differential equations for the general case is not considered in this paper.

ARMOUR INSTITUTE OF TECHNOLOGY.

THE PLATEAU PROBLEM FOR MINIMAL SURFACES OF ARBITRARY TOPOLOGICAL STRUCTURE.*

By Max Shiffman.

The classic problem of Plateau is to establish the existence of a minimal surface bounded by a given Jordan curve in space. Immediately after the solutions of this problem by J. Douglas and by T. Radó,[1] Douglas applied his method to prove the existence, under certain conditions, of a one-sided minimal surface bounded by a given curve and of a minimal surface bounded by two non-intersecting curves. He considered these as special cases of the following general form of the problem: to find a minimal surface having any prescribed topological structure and bounded by any finite number of given non-intersecting Jordan curves. In a note which appeared February, 1936, he stated his results concerning this general problem, and in a paper of June, 1936, he gave some details and proofs.[2] More recently, in August, 1938, Douglas published a series of notes in which he announced and summarized a forthcoming detailed publication.[2a]

In the meantime, in a note appearing June, 1936, R. Courant indicated a method for the solution of the general problem stated above, and also of the problem with free boundaries.[3] This method was presented in detail for the case of genus zero in a paper appearing in July, 1937. In this paper, Courant also points out that his method becomes somewhat simpler if conformal mapping theorems are used and that this procedure may be applied to the case of higher topological structure.

The purpose of the present paper, which is based on Courant's method,

* Received September 15, 1938.

[1] See Radó, "On the problem of Plateau," *Ergebnisse der Math.*, vol. 2 (1933), no. 2, for an excellent account of the Plateau problem and for further literature. A complete reference to Douglas's work is contained in Douglas, "The problem of Plateau," *Bulletin of the American Mathematical Society*, (1933), pp. 227-251.

[2] Douglas, "Some new results in the problem of Plateau" and "Minimal surfaces of general topological structure with any finite number of assigned boundaries," *Journal of Mathematics and Physics, M. I. T.*, vol. 15 (1936), pp. 55-64, 105-123; "Minimal surfaces of higher topological structure," *Proceedings of the National Academy of Sciences*, vol. 24 (1938), pp. 343-352, 353-359, 360-364.

[2a] This detailed publication has appeared: *Annals of Mathematics*, vol. 40 (1939), pp. 205-298, and *American Journal of Mathematics*, vol. 61 (1939), pp. 545-589, 590-608.

[3] Courant, *Proceedings of the National Academy of Sciences*, vol. 22 (1936), pp. 367-372, 373-375; Courant, "Plateau's problem and Dirichlet's principle," *Annals of Mathematics*, vol. 38 (1937), pp. 679-724. For the problem with free boundaries, see a note in *Proceedings of the National Academy of Sciences*, vol. 24 (1938), pp. 97-101.

is to carry out in detail the solution of the Plateau problem in the general form without making use of mapping theorems. This is of interest not only in the Plateau problem but also in obtaining normal forms of Riemann surfaces (although this is not carried through in this paper). This connection was emphasized by Douglas and developed by Courant in the paper quoted above for the case of genus zero.[3a]

In order that a solution to the general Plateau problem exist, certain restrictions must be made concerning the positions and shapes of the boundaries. The restrictions (in the form of inequalities) imposed by Douglas and Courant are such that the minimal surfaces obtained are absolute minima. The author has shown, in a previous paper,[4] that Courant's method can be generalized so as to include certain types of relative minima. In the present paper, a similar formulation is considered. A certain functional of a surface, called its "inner diameter," is introduced; it is a measure of how far the surface is from being of lower topological structure. In the problem of minimizing the Dirichlet functional, only those surfaces are admitted whose inner diameters lie between two positive bounds α, β. If the lower bound of the Dirichlet functional for all such surfaces is less than the lower bound for all surfaces whose inner diameters are exactly α or β, then it is shown that the variational problem has a solution and that this solution is the required minimal surface.

If one of the bounds α, β of the inner diameter is zero, we have the case of the absolute minimum. In this case, certain types of degenerations cannot be excluded. To cover this, the result must be formulated somewhat differently, or else the use of conformal mapping theorems is unavoidable. For this reason, this case is not discussed.

In § 2, the domains of representation of the surfaces are chosen as slit domains, and the topology of such domains is discussed. The limit of a sequence of slit domains is defined in § 3, and the necessary and sufficient condition that this limit domain have the given topological structure is obtained (Lemma 1). The inner diameter of a surface is defined in § 4, and the variational problem formulated. § 5 contains lemmas which are used throughout. The main theorem of this paper is stated on p. 864.

The variational problem is solved in §§ 6-8, making extensive use of §§ 3, 5. The solution is then identified as a minimal surface (§§ 9-13) by direct performance of the variations.

The author wishes to extend thanks to Professor Courant for his interest in this investigation.

[3a] Cf. also Douglas, *loc. cit.* note 2a, and a forthcoming paper by Bella Manel.

[4] Shiffman, "The problem of Plateau for minimal surfaces which are relative minima," *Annals of Mathematics*, vol. 39 (1938), pp. 309-315.

PRELIMINARIES.

1. Introduction. Let $\Gamma_1, \Gamma_2, \cdots, \Gamma_k$ be k prescribed non-intersecting closed Jordan curves in space. It is required to inscribe in them a minimal surface of prescribed topological structure, i. e., either an orientable or a non-orientable surface of characteristic q. In the case of orientability, the problem will be made more specific by requiring these orientations to coincide with given orientations of the curves $\Gamma_1, \cdots, \Gamma_k$. $\Gamma_1, \Gamma_2, \cdots, \Gamma_k$ are thus to be considered as oriented Jordan curves.

According to the Riemann-Weierstrass theorem, the surface

$$\mathfrak{x} = \mathfrak{x}(u, v)$$

(in vector terminology) in which u, v are isometric parameters, i. e., $E = G$, $F = 0$ where E, F, G are the first fundamental quantities of the surface, $E = \mathfrak{x}_u^2$, $F = \mathfrak{x}_u\mathfrak{x}_v$, $G = \mathfrak{x}_v^2$, is a minimal surface if and only if

$$\mathfrak{x}_{uu} + \mathfrak{x}_{vv} = 0.$$

The potential character of $\mathfrak{x}(u, v)$ implies that the following function

$$\phi(w) = (\mathfrak{x}_u - i\mathfrak{x}_v)^2 = E - G - 2iF$$

is an analytic function of $w = u + iv$, and the isometric character of the parameters u, v that $\phi(w) \equiv 0$.

Our starting point for the solution of the Plateau problem is to minimize the Dirichlet functional

$$D(\mathfrak{x}) = \tfrac{1}{2} \int\int (E + G)\,dudv = \tfrac{1}{2} \int\int (\mathfrak{x}_u^2 + \mathfrak{x}_v^2)\,dudv$$

among surfaces $\mathfrak{x} = \mathfrak{x}(u, v)$ of the prescribed topological structure bounded by the given curves.

The idea of minimizing the Dirichlet functional for the solution of the Plateau problem was introduced by Douglas, who uses potential surfaces as the admissible functions. He obtains an expression, which he calls $A(g)$, of the Dirichlet functional of a potential surface in terms of its boundary values, and then considers the problem of minimizing $A(g)$. Courant starts directly with the problem of minimizing $D(\mathfrak{x})$ among all surfaces, without restriction to potential surfaces. This avoids many of the complications of the Douglas procedure, and makes accessible many other problems (e. g., the Plateau problem with free boundaries). It is Courant's procedure which we adopt.

2. Domain of representation. The parameters u, v in the surface $\mathfrak{x} = \mathfrak{x}(u, v)$ must range over a manifold of higher topological structure. Since the theory of conformal mapping of Riemann surfaces is contained in the

Plateau problem, this manifold should be chosen as a normal domain of a Riemann surface. Such normal domains are well-known, the type which we shall use being a modified form of the "slit" domains developed by Hilbert, Courant, Koebe.[5]

The domain of representation will be taken, in general, as the half (u, v)-plane above the u-axis in which rectilinear cuts parallel to the u-axis are made. There are $k - 1$ finite cuts which, together with the u-axis, correspond to the

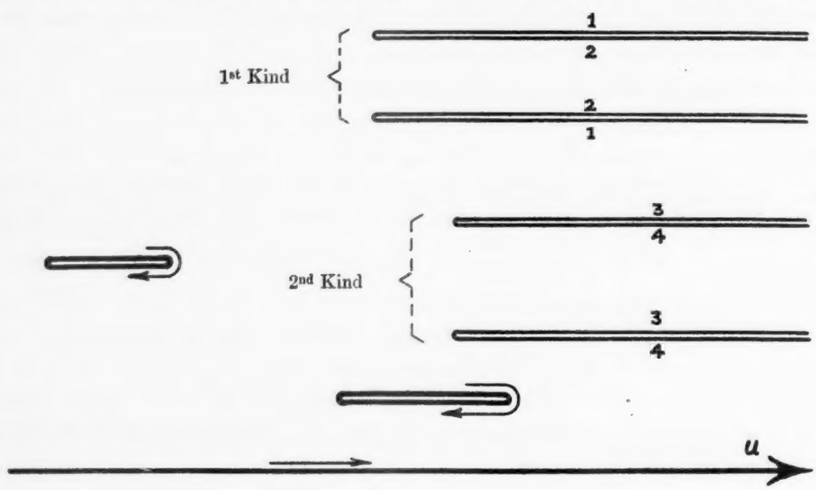


Fig. 1.

boundaries (the arrows in the diagram signify the orientations of these boundary slits); σ pairs of infinite cuts (extending to infinity to the right) of the first kind and τ pairs of infinite cuts of the second kind, with edges coördinated as in the diagram. Those edges marked with the same numeral are to be joined by identifying points which are directly below or above one another. The characteristic q is to be $k + \sigma + \tau - 2$, and the domain orientable or non-orientable according as $\tau = 0$ or $\tau > 0$.[6]

[5] Hilbert, *Gott. Nachr.* (1909), pp. 314-323; Courant, *Mathematische Zeitschrift*, vol. 3 (1919), pp. 114-122; Koebe, *Gott. Nachr.* (1919), pp. 1-46.

[6] The fact that any Riemann surface of finite genus can be mapped conformally into a slit domain of the type just described is easily proved on the basis of the proof of the normal domains of Hilbert, Courant, Koebe. The Riemann surface can be mapped conformally into a domain of the plane in which one boundary is a circle. Double the surface across the boundary by reflection, and construct the dipole potential with singularity at any point of the circle and axis perpendicular to the circle. The corresponding analytic function maps the original domain into a slit domain of the type described.

In the case of a non-orientable Riemann surface, consider the double of the surface (with respect to orientation) and construct the potential function with the same dipole

The slits must suitably interlock one another in order that the characteristic of the domain be equal to $k + \sigma + \tau - 2$. The criterion for this is the following. Let u_0 be a sufficiently large abscissa such that all the infinite slits and none of the boundary slits lie above u_0. Beginning from the u-axis at $u = u_0$, draw the vertical line in the domain such that if it meets a slit it is continued from the coördinated slit. The domain has the characteristic $k + \sigma + \tau - 2$ if and only if this vertical curve fills out the whole line $u = u_0$, $v \geqq 0$, i. e., if the point at infinity has a complete neighborhood surrounding it.[7] For, mark the slits of the domain and count the number c_2 of faces, c_1 of edges, and c_0 of vertices (omit the k boundaries, which contribute k to the characteristic). In the count of vertices, the infinite counts as $1 + \gamma$ points where $\gamma = 0$ or $\gamma > 0$ according as the vertical curve fills out the whole line $u = u_0$ or not. We have $c_2 = 1$, $c_1 = 2\sigma + 2\tau$, $c_0 = \sigma + \tau + 1 + \gamma$. By the Euler formula,[8] taking into account the k boundaries, the characteristic is $k + (-c_2 + c_1 - c_0) = k + \sigma + \tau - 2 - \gamma$ which is equal to $k + \sigma + \tau - 2$ if and only if $\gamma = 0$.

It is necessary to admit the case when some of the slits have coalesced.[9] The complete definition of the slit domains is then as follows. A slit domain consists of the upper half (u, v)-plane above the u-axis in which finite and infinite slits (extending to infinity toward the right) parallel to the u-axis are made. Each side of each slit, and the u-axis, is divided into a finite number of pieces; each piece is either a part of the boundary of the domain, or is coördinated to exactly one other such piece lying directly below or above it by identifying points with the same abscissa (they may be on opposite sides of one slit). It is required, in addition, that the point at infinity have a complete neighborhood surrounding it.

A single boundary of a slit domain consists of all those boundary pieces which are connected to each other by the coördinations of the slits. The topological requirements of the slit domain are that there be k boundaries, that the characteristic of the domain be equal to q (this condition will be put in a more convenient form below), and that all pairs of coördinated pieces occur

---

singularity at a pair of symmetrical points. The corresponding analytic function maps the original non-orientable surface into a slit domain with infinite slits of the second kind as well as of the first kind. Infinite slits of the second kind correspond to stream lines, through crossing points, which pass through both singularities. See a paper by Courant soon to be published.

[7] In obtaining the mapping of a Riemann surface on a slit domain, this condition is automatically fulfilled, and the structure of the slits then determined.

[8] For the Euler formula and its use without triangulation, see Seifert-Threlfall, *Lehrbuch der Topologie*, esp. pp. 145, 146.

[9] Numerous examples of coalesced slit domains are given by Koebe, Courant, *loc. cit.*, note 5.

on different sides of slits (i. e., one on the upper side of a slit and the other on the lower side of a slit), or not, according as the domain is to be orientable or non-orientable respectively.

Finally, the slit domains will be normalized. The infinite slit which extends most to the left will extend to $u = 0$, and the topmost infinite slit will lie on the line $v = 1$. If the characteristic is $k - 2$ (no infinite slits, genus zero), the center of one of the boundary slits will be at $(0, 1)$.[10]

The following considerations will be limited to the case when all the slits are infinite, since any finite slit can be extended to infinity, opposite sides of the extension being coördinated. A pair of coördinated pieces, or a boundary piece, is called an edge of the slit domain; a point which separates edges, together with all its coördinated points, is a vertex of the slit domain. A vertex is composed of several points, identified; let a be the number of these points which occur on the extreme left ends of infinite slits, and b the number which do not. Around each of the a endpoints there is an angle 2π; around the b other points, an angle π. To this vertex, therefore, correspond $a + b$ points, identified, and an angle of $2\pi(a + b/2)$. If the vertex does not lie on a boundary of the slit domain, it is called an inner *branch-point* of order $\nu = a + b/2 - 1$; if the vertex lies on a boundary, it is a boundary *branch-point* of order $\mu = a + b/2 - \frac{1}{2}$.[11]

For an inner branch-point, b must be even; for a boundary branch-point, b may be odd or even. At a boundary branch-point, there are exactly two boundary edges. If the order is an integer (b odd), these two boundary edges issue in opposite directions from the vertex; if not (b even), they issue in the same direction from the vertex. On a single boundary, there are just as many branch points with both boundary edges issuing to the right as there are branch points with both boundary edges issuing to the left.

Let e be the number of edges of the slit domain, v the number of vertices (not including the point at infinity). By the Euler formula, the characteristic q of the domain is equal to $-1 + e - (v + 1)$, so that $e - v = q + 2$. This condition will be put in another form, namely, that the sum ρ of the orders of all the branch points of the slit domain is equal to $q + 1$.[12] Let $n_{a,b}$ be the number of inner vertices with a extreme end-points of slits and b others, $m_{a,b}$ the number of such boundary vertices, s the number of slits (exclusive

[10] This normalization fixes 5 of the free parameters in the mapping of a Riemann surface into a slit domain. The additional parameter corresponds to that point of the boundary chosen as the infinite point of the slit domain.

[11] This is one-half the order of the resulting inner branch point when the domain is doubled across the boundary.

[12] If there are no boundaries, $\rho = q + 2$.

of the u-axis), e_1 the number of boundary edges, e_2 the number of inner edges. Then

$$\rho = \sum_{a,b} n_{a,b}(a + b/2 - 1) + \sum_{a,b} m_{a,b}(a + b/2 - \tfrac{1}{2}),$$

$$s = \sum_{a,b} (n_{a,b} + m_{a,b})a, \qquad v = \sum_{a,b} (n_{a,b} + m_{a,b}).$$

The number of boundary vertices is $\sum_{a,b} m_{a,b}$; therefore the number of boundary edges is $e_1 = 1 + \sum_{a,b} m_{a,b}$, the additional 1 arising because of the u-axis. The

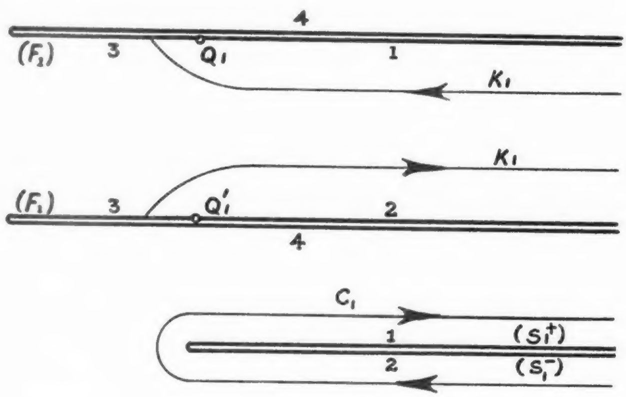


Fig. 2.

total number of points making up all the vertices is $\sum_{a,b} (n_{a,b} + m_{a,b})(a + b)$; therefore,

$$2e_2 + e_1 = \sum_{a,b} (n_{a,b} + m_{a,b})(a + b) + (s + 1),$$

or

$$e_2 = \sum_{a,b} n_{a,b}(a + b/2) + \sum_{a,b} m_{a,b}(a + b/2 - \tfrac{1}{2}).$$

Hence

$$e - v = e_1 + e_2 - v = \sum_{a,b} n_{a,b}(a + b/2 - 1) + \sum_{a,b} m_{a,b}(a + b/2 - \tfrac{1}{2}) + 1$$
$$= \rho + 1.$$

Thus,[13]

$$\rho = q + 1.$$

We shall now obtain a complete homology basis of curves in a slit domain. Consider an inner branch point of order ν with a extreme end-points $P_1, \cdots, P_a$,

[13] This is equivalent to the statement that the difference between the number of zeros and the number of poles of a single valued differential on a closed Riemann surface of genus p is $2p - 2$. Cf. Weyl, *Die Idee der Riemannschen Flächen*, esp. pp. 124, 125.

and $2c$ others $Q_1, Q'_1, \cdots, Q_c, Q'_c$ where Q_i, Q'_i are paired so that the edges immediately to their left are coördinated, and $\nu = a + c - 1$. Let S_j be the slit with P_j as end-point, and E_i, F_i and E'_i, F_i the edges with Q_i and Q'_i as common end-points respectively. Let C_j be a curve surrounding S_j, and K_i a curve surrounding E_i ending in a point of F_i, and beginning from the coördinated point of F_i surrounding E'_i. These $\nu + 1$ curves $C_1, \cdots, C_a$, $K_1, \cdots, K_c$ all pass through the point at infinity. Take any ν of them as part of the desired homology basis.

Consider, next, a boundary branch point of integral order μ with a extreme end-points $P_1, \cdots, P_a$ and $2c + 1$ others, $Q_1, Q'_1, \cdots, Q_c, Q'_c, Q$ where Q_i, Q'_i are paired as above, Q is the point which has a boundary edge to its left, and $\mu = a + c$. Construct the μ curves $C_1, \cdots, C_a$, $K_1, \cdots, K_c$ as above. These will be taken as part of the homology basis. Consider, finally, a boundary branch point of non-integral order μ with a extreme end-points $P_1, \cdots, P_a$ and $2c$ others $Q_1, Q'_1, \cdots, Q_c, Q'_c$ paired as previously, where $\mu = a + c - \frac{1}{2}$. If the two boundary edges issue to the right, the $\mu + \frac{1}{2}$ curves $C_1, \cdots, C_a$, $K_1, \cdots, K_c$ constructed as above will be part of the homology basis. If the two boundary edges issue to the left, let Q_c, Q'_c be the pair of points which have these boundary edges to their left; the $\mu - \frac{1}{2}$ curves $C_1, \cdots, C_a$, $K_1, \cdots, K_{c-1}$ will then be taken as part of the homology basis.

Proceeding in this way for all the branch points of the domain, one obtains a total of ρ curves. To each such curve corresponds the vertex about which it is drawn. We shall show that no combination $S_1, \cdots, S_s$ of these curves can be the complete boundary of a part of the slit domain. Consider those curves of $S_1, \cdots, S_s$ whose vertices have the smallest abscissa, and let $T_1, \cdots, T_t$ be the curves among these corresponding to a single vertex V. If one of $T_1, \cdots, T_t$ contains a boundary edge issuing from V, $S_1, \cdots, S_s$ could not be the complete boundary of a part of the slit domain. If none of $T_1, \cdots, T_t$ contains a boundary edge issuing from V, a side of T_1 can be connected to the opposite side by a path which does not cross any of the curves $S_1, \cdots, S_s$. It follows that no combination $S_1, \cdots, S_s$ of the ρ curves can be the complete boundary of a part of the slit domain. Since $\rho = q + 1$, these ρ curves form a complete homology basis of the slit domain, by definition of the characteristic q. This is the desired homology basis.

A curve in a slit domain $\mathfrak{G}$ which is closed or joins boundary points of $\mathfrak{G}$ will be called relatively closed. A relatively closed curve will be called non-bounding if it is one of the following three types: a closed curve non-bounding in $\mathfrak{G}$, a curve joining different boundaries of $\mathfrak{G}$, or a curve joining points of one boundary but forming together with the arcs of this boundary closed curves non-bounding in $\mathfrak{G}$. The minimum of the Euclidean lengths of all

such non-bounding relatively closed curves in $\mathfrak{G}$ is called the ***inner diameter*** of $\mathfrak{G}$. The inner diameter is a measure of how far the slit domain is from being of lower topological structure.

A relatively closed curve in a slit domain which has at most one point in common with any horizontal line is non-bounding, since both sides of the curve can be joined to the point at infinity.

A metric can be defined for each slit domain $\mathfrak{G}$ as follows: the distance $\overline{PQ}$ between two points P, Q of $\mathfrak{G}$ is the smallest of the Euclidean lengths of curves joining P, Q in $\mathfrak{G}$. It is easily seen that the triangle inequality holds and that $\overline{PQ} = 0$ if and only if P, Q are the same point in $\mathfrak{G}$.

3. Sequences of slit domains. Let $\mathfrak{G}_1, \mathfrak{G}_2, \cdots, \mathfrak{G}_n, \cdots$ be a sequence of slit domains with k boundaries and characteristic q such that the number of slits in $\mathfrak{G}_n$ and their relative situations [14] are the same for all n, and corresponding vertices converge to finite points. We shall define, by continuity, a limit slit domain $\mathfrak{G}$.

The slits of $\mathfrak{G}$ are the limits of the slits of $\mathfrak{G}_n$ as $n \to \infty$. Let $u_1, u_2, \cdots, u_h$, $u_1 < u_2 < \cdots < u_h < \infty$, be the limits of the abscissae of the vertices of $\mathfrak{G}_n$ as $n \to \infty$; the vertices of $\mathfrak{G}$ will have $u_1, u_2, \cdots, u_h$ as abscissae. The coördinations in $\mathfrak{G}$ are as follows. Let P be a point with abscissa u different from $u_1, u_2, \cdots, u_h$ on the upper edge (lower edge) of a slit S of $\mathfrak{G}$. For all sufficiently large n, the slits of $\mathfrak{G}_n$ above u are similarly arranged and coördinated. Consider all those slits of $\mathfrak{G}_n$ which tend to the slit S of $\mathfrak{G}$ as $n \to \infty$, and let P_n be the point in $\mathfrak{G}_n$ above u on the upper edge of the highest of these slits (on the lower edge of the lowest of these slits). P will be taken as a boundary point if and only if the distance in $\mathfrak{G}_n$ of P_n from a boundary of $\mathfrak{G}_n$ tends to zero as $n \to \infty$. Let P' be another point above u on a slit S' of $\mathfrak{G}$, and P'_n the corresponding point in $\mathfrak{G}_n$. Then P and P' will be coördinated if and only if the distance $\overline{P_nP'_n}$ in $\mathfrak{G}_n \to 0$ as $n \to \infty$. Thus, every point above u on a slit in $\mathfrak{G}$ is either a boundary point or is coördinated to exactly one other such point. This description is the same for every u in each interval $u_i < u < u_{i+1}$, $i = 1, \cdots, h-1$, $u_h < u < \infty$; for the abscissae u_i, $i = 1, \cdots, h$, it is determined by that around u_i. This completes the definition of the slit domain $\mathfrak{G}$. It is easily seen, from the same property for $\mathfrak{G}_n$, that the point at infinity in $\mathfrak{G}$ has a complete neighborhood surrounding it.

The limit $\mathfrak{G}$ may likewise be constructed as follows (see Fig. 3 below). By introducing new vertices (of order 0) in $\mathfrak{G}_n$ if necessary, we may suppose

---

[14] By the relative situation of the slits is meant the arrangement of the boundary edges and the coördinated edges in the order of increasing abscissae and ordinates of their end points.

that each edge of $\mathfrak{G}_n$ whose length does not tend to zero lies above an interval of the u-axis which tends to one of $u_i < u < u_{i+1}$, $i = 1, \cdots, h-1$, $u_h < u < \infty$, as $n \to \infty$. In the neighborhood of u_i, consider all those slits of $\mathfrak{G}_n$ which tend to one slit of $\mathfrak{G}$; successively join the sides of these slits by straight lines in the neighborhood of u_i. Do this for all slits of $\mathfrak{G}$ and all u_i. This decomposes $\mathfrak{G}_n$ into a large domain and several thin strips. If a horizontal edge of one of the thin strips is coördinated to a horizontal edge of the large domain, attach the strip to the large domain along this edge on the side opposite to the large domain. Continue until either no strips remain or no edge of any remaining strip is coördinated to any horizontal edge of the large domain. The final large domain $\bar{\mathfrak{G}}_n$ will have horizontal edges and small additional edges joining adjacent horizontal edges. From the definition of $\mathfrak{G}$, it is immediately seen that the horizontal edges of $\bar{\mathfrak{G}}_n$ are coördinated just as in $\mathfrak{G}$, so that the limit of $\bar{\mathfrak{G}}_n$ as $n \to \infty$ is $\mathfrak{G}$.[15]

We shall now determine the condition that $\mathfrak{G}$ have the same topological structure as every $\mathfrak{G}_n$. It is contained in

LEMMA 1. *The limit domain $\mathfrak{G}$ has the same topological structure as every $\mathfrak{G}_n$ if and only if the inner diameter*[16] *t_n of $\mathfrak{G}_n$ does not tend to zero as $n \to \infty$.*

Proof. Construct the domain $\bar{\mathfrak{G}}_n$ as above. Around each vertex V_i of $\mathfrak{G}$ construct a complete neighborhood N_i in $\mathfrak{G}$ bounded by a relatively closed curve J_i of length L. In $\bar{\mathfrak{G}}_n$ these same curves $J_i^{(n)}$ contain in their interiors $N_i^{(n)}$ small irregular edges of $\bar{\mathfrak{G}}_n$. $\mathfrak{G}$ is decomposed into a large domain $\mathfrak{G}^*$ and all the simply connected pieces N_i; and $\bar{\mathfrak{G}}_n$ is decomposed into the same $\mathfrak{G}^*$ and all the $N_i^{(n)}$.

If $t_n \to t \neq 0$, in the construction of $\bar{\mathfrak{G}}_n$ from $\mathfrak{G}_n$ there are no remaining pieces. Otherwise, by attaching remaining strips together along coördinated horizontal edges to form a cycle (either closed or bounded by boundary edges), there would be non-bounding relatively closed curves in $\mathfrak{G}_n$ of length tending to zero as $n \to \infty$. Hence $\mathfrak{G}_n$ and $\bar{\mathfrak{G}}_n$ are topologically equivalent. Now, choose L less than every t_n. Then each relatively closed curve $J_i^{(n)}$, having a length $< t_n$, must bound in $\bar{\mathfrak{G}}_n$. Their interiors $N_i^{(n)}$ in $\bar{\mathfrak{G}}_n$ are therefore distinct pieces. These pieces are simply connected; otherwise there would be non-bounding relatively closed curve in $N_i^{(n)}$ whose length is less than that of

[15] It is easily established that a coalesced slit domain is the limit of a sequence of slit domains of the same topological structure in which the slits are completely separated if and only if each boundary is intersected by every vertical line in at most two points.

[16] See section 2, p. 861, for the definition of the inner diameter of a slit domain.

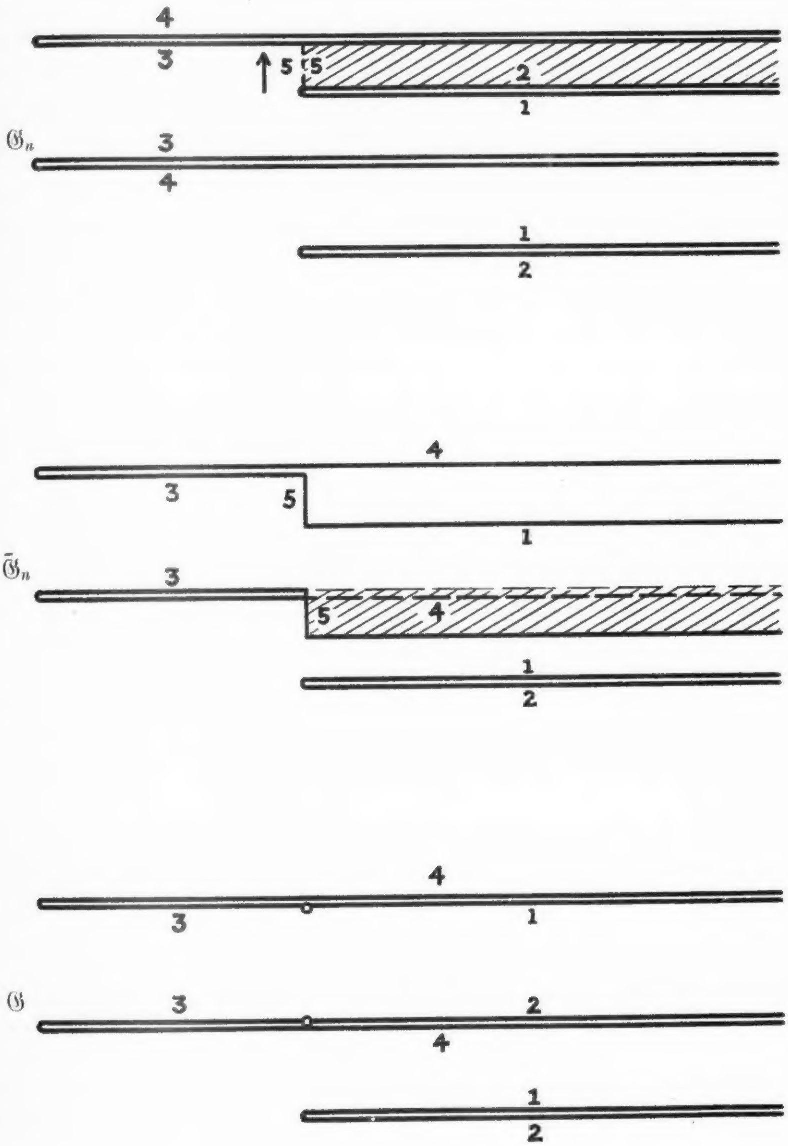

4
3
5
5
2
1
$\mathfrak{G}_n$
3
4
1
2
4
3
5
1
$\bar{\mathfrak{G}}_n$
3
5
4
1
2
4
3
1
$\mathfrak{G}$
3
2
4
1
2

Fig. 3.

$J_i^{(n)}$, or less than t_n. The domains $\mathfrak{G}$ and $\bar{\mathfrak{G}}_n$ are therefore topologically equivalent.

If $t_n \to 0$, let L have any small value. Since there must be at least one non-bounding relatively closed curve in $\mathfrak{G}_n$ of length tending to zero as $n \to \infty$, either $\bar{\mathfrak{G}}_n$ is not the complete $\mathfrak{G}_n$ (i. e., strips have remained), or the $N_i^{(n)}$ are not distinct simply connected regions, or both. In any case, the characteristic of $\mathfrak{G}$ is less than that of $\mathfrak{G}_n$.

4. Inner diameter of a surface. The main theorem and the minimum problem. Let $\mathfrak{x} = \mathfrak{x}(u, v)$ be a surface in $\mathfrak{x}$-space, the parameters u, v ranging over a slit domain $\mathfrak{G}$ of the (u, v)-plane. Let C be a non-bounding closed curve in $\mathfrak{G}$, and let t_c be the oscillation of $\mathfrak{x}(u, v)$ on C. Define the *inner diameter* of the surface $\mathfrak{x}(u, v)$ as the greatest lower bound of t_c for all such closed curves C non-bounding in $\mathfrak{G}$.

It is clear that if the surfaces $\mathfrak{x}_n(u, v)$ converge uniformly to the surface $\mathfrak{x}(u, v)$, the inner diameters t_n of $\mathfrak{x}_n$ tend to the inner diameter t of $\mathfrak{x}$.

Consider surfaces $\mathfrak{x} = \mathfrak{x}(u, v)$ which are continuous and have piecewise continuous first derivatives [17] over a slit domain $\mathfrak{G}$ of the (u, v)-plane of the prescribed topological structure (with k boundaries, of characteristic q, and either orientable or non-orientable), and which map the boundaries on the curves $\Gamma_1, \Gamma_2, \cdots, \Gamma_k$, preserving orientation in the case of orientability. We call these "allowable" surfaces. Consider the Dirichlet integral

$$D(\mathfrak{x}) = \tfrac{1}{2} \iint\limits_{\mathfrak{G}} (\mathfrak{x}_u^2 + \mathfrak{x}_v^2)\, du dv.$$

Let d_α be the greatest lower bound of $D(\mathfrak{x})$ for all allowable surfaces whose inner diameters are exactly equal to α. Let $d_{\alpha,\beta}$ be the greatest lower bound of $D(\mathfrak{x})$ for all allowable surfaces whose inner diameters lie between α and β, inclusive. We assert the

MAIN THEOREM 1. *If, for two positive numbers α, β $(0 < \alpha < \beta)$,*

$$d_{\alpha,\beta} < d_\alpha, \qquad d_{\alpha,\beta} < d_\beta$$

then there exists a minimal surface of the prescribed topological structure bounded by the curves $\Gamma_1, \Gamma_2, \cdots, \Gamma_k$ with inner diameter between α and β.[18]

The proof is based on the following minimum problem:

$$D(\mathfrak{x}) = \text{minimum}$$

[17] Hereafter, all surfaces will be considered as continuous and having piecewise continuous first derivatives.

[18] The inequalities require, in particular, that $d_{\alpha,\beta}$ be finite.

among all allowable surfaces $\mathfrak{x}(u, v)$ whose inner diameters lie between α and β, inclusive. The fact that α, β are positive (not zero) is essential for the proof.

5. Preliminary lemmas. We shall require several lemmas similar to lemmas of Courant.

LEMMA 2. *Let $\mathfrak{x}(u, v)$ be a surface defined over a region H bounded by two concentric squares with sides of length $2a$ and $2b$ $(a < b)$, and by two curves joining these squares. Let t_ρ be the oscillation of $\mathfrak{x}(u, v)$ on that part Q_ρ in H of the concentric square of side 2ρ, $a \leqq \rho \leqq b$, and let t be the minimum of t_ρ. Then*

$$\frac{t^2}{64} \log \frac{b}{a} \leqq D(\mathfrak{x}).$$

Proof. Let τ be any number less than t. The oscillation of $\mathfrak{x}(u, v)$ on at least one side of Q_ρ is $> \tau/4$. In the interval $c \leqq \rho \leqq d$, let this side be on the line $u = \rho$, where (u, v) is a coördinate system with origin at the center of the concentric squares and axes parallel to the sides of the squares. Then, along this line,

$$\frac{\tau}{4} < \left| \int_A^B \mathfrak{x}_v dv \right| \leqq \int_A^B | \mathfrak{x}_v | \, dv \leqq \int_C^D | \mathfrak{x}_v | \, dv,$$

where A, B are points on the line $u = \rho$ for which $| \mathfrak{x}(B) - \mathfrak{x}(A) | > \tau/4$, and C, D are the end-points of the line in H. By Schwarz's inequality,

$$\frac{\tau^2}{16} \leqq 2\rho \int_C^D \mathfrak{x}_v^2 dv.$$

Divide by $4u$, noting that $\rho = u$, and integrate with respect to u between the limits c, d:

$$\frac{\tau^2}{64} \log \frac{d}{c} \leqq \tfrac{1}{2} \int_c^d \int_C^D \mathfrak{x}_v^2 dv \leqq \tfrac{1}{2} \int_c^d \int_C^D (\mathfrak{x}_u^2 + \mathfrak{x}_v^2) dudv.$$

Similarly, if the side on which the oscillation of $\mathfrak{x}(u, v)$ is $> \tau/4$ is any one of the sides. The intervals $c \leqq \rho \leqq d$ fill out the whole interval $a \leqq \rho \leqq b$. Adding all the corresponding inequalities, we obtain

$$\frac{\tau^2}{64} \log \frac{b}{a} \leqq D(\mathfrak{x}).$$

This holding for all $\tau < t$, it holds for $\tau = t$.

LEMMA 3. *Let $\mathfrak{x}(u, v)$ be a surface defined over a rectangular domain $0 \leqq u \leqq L$, $0 \leqq v \leqq h$ of length L and height h, and t the minimum oscillation of $\mathfrak{x}(u, v)$ on vertical lines $u = constant$. Then*

$$\frac{t^2L}{2h} \leqq D(\mathfrak{x}).$$

Proof. As in the proof of Lemma 2,

$$t^2 \leqq h \int_0^h \mathfrak{x}_v{}^2 dv \quad \text{and} \quad t^2L \leqq 2hD(\mathfrak{x}).$$

LEMMA 4. *Let $p(u,v)$ be a potential function defined in a domain bounded partly by the line $v=0$, $a \leqq u \leqq b$. Suppose that*

$$\lim_{\delta\to 0} \int_{a'}^{b'} \lambda(u,\delta)p(u,\delta)\,du = 0 \qquad (a \leqq a' < b' \leqq b)$$

for any arbitrary function $\lambda(u,v)$ with continuous and bounded first derivatives for which $\lambda(a',v) = \lambda(b',v) = 0$. Then $p(u,v)$ has the boundary values 0 on $v=0$, $a' < u < b'$.

Proof. Choose a rectangle R, $a' \leqq u \leqq b'$, $0 \leqq v \leqq c$, which lies entirely in the domain, and let (U,V) be any point in it. Consider the rectangle R_δ, $a' \leqq u \leqq b'$, $\delta \leqq v \leqq c+\delta$, for δ sufficiently small, and let $g_\delta(u,v;\,U,V)$ be the Green function for R_δ with singularity at (U,V). Let $q_\delta(u,v)$ be the bounded potential function in R_δ whose boundary values are 0 on $v=\delta$ and equal to $p(u,v)$ on the remaining boundary lines. Then

$$p(U,V) - q_\delta(U,V) = \int_{a'}^{b'} \frac{\partial g_\delta(u,\delta;\,U,V)}{\partial v}\, p(u,\delta)\,du.$$

Now, by virtue of the congruence of R and R_δ,

$$g_\delta(u,v;\,U,V) = g(u,v-\delta;\,U,V-\delta)$$

and

$$\frac{\partial g_\delta(u,\delta;\,U,V)}{\partial v} = \frac{\partial g(u,0;\,U,V-\delta)}{\partial v},$$

where $g(u,v;\,U,V)$ is the Green function for R. This shows that

$$\frac{\partial g_\delta(u,\delta;\,U,V)}{\partial v},$$

considered as a function of u and δ, has continuous and bounded derivatives in the neighborhood of $\delta=0$. It also vanishes for $u=a'$, $u=b'$ since $g_\delta(u,v;\,U,V)=0$ on $u=a'$ and $u=b'$. By the hypothesis of the lemma, choosing

$$\lambda(u,\delta) = \frac{\partial g_\delta(u,\delta;\,U,V)}{\partial v}$$

in the neighborhood of $\delta=0$, the integral above tends to zero as $\delta\to 0$, so that $q_\delta(U,V) \to p(U,V)$. But $q_\delta(u,v)$ can be extended by reflection across

the line $v = \delta$, and it converges to a potential function whose values on $v = 0$ are 0. Hence $p(u, v)$ has the boundary values 0 on $v = 0$, $a' < u < b'$.

EXISTENCE OF A SOLUTION TO THE MINIMUM PROBLEM.

6. Minimizing sequences. The problem is to minimize $D(\mathfrak{x})$ among all allowable surfaces whose inner diameters lie between α and β. Let

$$\mathfrak{x}_1, \mathfrak{x}_2, \cdots, \mathfrak{x}_n, \cdots$$

be a minimizing sequence, $D(\mathfrak{x}_n) \rightarrow d_{\alpha,\beta}$. Because of the inequalities $d_{\alpha,\beta} < d_\alpha$, $d_{\alpha,\beta} < d_\beta$, we may suppose that $D(\mathfrak{x}_n) < d_\alpha, d_\beta$ for all n. The sequence $\mathfrak{x}_n$ will be replaced by a new minimizing sequence consisting solely of potential surfaces.

Let $\mathfrak{x}(u, v)$ be a fixed member $(= \mathfrak{x}_n)$ of the minimizing sequence, and $\bar{\mathfrak{x}}(u, v)$ the bounded potential surface with the same boundary values as $\mathfrak{x}(u, v)$.[19] Define the surface $\mathfrak{x}_\epsilon(u, v)$, for $0 \leq \epsilon \leq 1$, by

$$\mathfrak{x}_\epsilon(u, v) = \bar{\mathfrak{x}}(u, v) + \epsilon\zeta$$

where $\zeta = \mathfrak{x}(u, v) - \bar{\mathfrak{x}}(u, v)$ and ζ has the boundary values zero. This is a continuous family of surfaces for which

$$\mathfrak{x}_1(u, v) = \mathfrak{x}, \qquad \mathfrak{x}_0(u, v) = \bar{\mathfrak{x}}.$$

The inner diameter of $\mathfrak{x}_\epsilon(u, v)$ is therefore a continuous function of ϵ. By the minimizing character of potential surfaces with regard to the Dirichlet functional,[20]

$$D(\mathfrak{x}_\epsilon) = D(\bar{\mathfrak{x}}) + \epsilon^2 D(\zeta) = D(\mathfrak{x}) - (1 - \epsilon^2) D(\zeta) \leq D(\mathfrak{x}).$$

The inner diameter of $\mathfrak{x}_\epsilon$ can never take the values α or β since $D(\mathfrak{x}_\epsilon) \leq D(\mathfrak{x})$

---

[19] In defining a potential function across a pair of coördinated edges, one must have a coördination of directions as well. The positive u directions at both edges are coördinated; the positive v direction at one edge is coördinated to the positive or negative v direction at the other edge according as the edges are of the first kind or second kind respectively.

[20] Among all surfaces with given boundary values, the potential surface $\bar{\mathfrak{x}}$ has the smallest Dirichlet integral. If ζ is any surface with boundary values 0 and with a finite Dirichlet integral, and if $\mathfrak{x}_\epsilon = \bar{\mathfrak{x}} + \epsilon\zeta$, then $D(\mathfrak{x}_\epsilon) \geq D(\bar{\mathfrak{x}})$. Since

$$D(\mathfrak{x}_\epsilon) = D(\bar{\mathfrak{x}}) + 2\epsilon D(\bar{\mathfrak{x}}, \zeta) + \epsilon^2 D(\zeta),$$

the usual argument yields $D(\bar{\mathfrak{x}}, \zeta) = 0$, so that

$$D(\mathfrak{x}_\epsilon) = D(\bar{\mathfrak{x}}) + \epsilon^2 D(\zeta).$$

This applies to the case $\zeta = \mathfrak{x} - \bar{\mathfrak{x}}$; for then

$$D(\zeta) \leqq (\sqrt{D(\mathfrak{x})} + \sqrt{D(\bar{\mathfrak{x}})})^2 \leq 4D(\mathfrak{x}).$$

$< d_\alpha, d_\beta$. Since the inner diameter of $\mathfrak{x}(u,v)$ lies between α and β, it follows that the inner diameter of $\bar{\mathfrak{x}}(u,v)$ lies between α and β.

Replace $\mathfrak{x}(u,v)$ by the potential surface $\bar{\mathfrak{x}}(u,v)$ for each member of the minimizing sequence. This is a minimizing sequence, which we rename $\mathfrak{x}_1, \mathfrak{x}_2, \cdots, \mathfrak{x}_n, \cdots$, consisting of potential surfaces.

7. Convergence of the domains. Let $\mathfrak{G}_1, \mathfrak{G}_2, \cdots, \mathfrak{G}_n, \cdots$ be the slit domains, normalized according to section 2, corresponding to the surfaces $\mathfrak{x}_1, \mathfrak{x}_2, \cdots, \mathfrak{x}_n, \cdots$ of the minimizing sequence. We may suppose, by choosing a subsequence if necessary, that the slits in $\mathfrak{G}_n$ are similarly situated for all n and that the vertices of $\mathfrak{G}_n$ converge to points either in the finite or at infinity as $n \to \infty$. We shall show that no vertices can tend to infinity and that the inner diameter of the domain $\mathfrak{G}_n$ cannot tend to zero as $n \to \infty$. The domains $\mathfrak{G}_n$ therefore converge to a slit domain $\mathfrak{G}$ of the prescribed topological structure.

Let δ be the minimum distance between the curves $\Gamma_1, \Gamma_2, \cdots, \Gamma_k$. We shall first prove the

LEMMA 5. *The oscillation of $\mathfrak{x}_n$ on a vertical curve V_n in $\mathfrak{G}_n$ which is relatively closed cannot tend to zero as $n \to \infty$.*

Proof. If V_n is a closed curve, it is non-bounding in $\mathfrak{G}_n$ and the oscillation of $\mathfrak{x}_n$ on V_n must be $\geqq \alpha$. If V_n joins different boundaries of $\mathfrak{G}_n$, the oscillation of $\mathfrak{x}_n$ on V_n must be $\geqq \delta$. The remaining possibility is that V_n join points of one boundary S_n. Let P_n, P'_n be the end-points of V_n on S_n, and γ_n, γ'_n the two arcs into which S_n is divided by P_n, P'_n. If the oscillation of $\mathfrak{x}_n$ on V_n tends to zero, $|\mathfrak{x}_n(P_n) - \mathfrak{x}_n(P'_n)| \to 0$ in particular. Since $\mathfrak{x}_n$ maps S_n monotonically onto one of the Jordan curves $\Gamma_1, \Gamma_2, \cdots, \Gamma_k$, the oscillation of $\mathfrak{x}_n$ on at least one of the two arcs γ_n, γ'_n tends to zero. This arc, combined with V_n, forms a closed curve C_n on which the oscillation of $\mathfrak{x}_n$ tends to zero. But C_n is non-bounding in $\mathfrak{G}_n$, and the oscillation of $\mathfrak{x}_n$ on C_n must be $\geqq \alpha$. This is a contradiction, and the lemma is proved.

(a) *No boundaries or vertices of infinite slits can tend to the point at infinity.* Consider a boundary S_n of $\mathfrak{G}_n$, and let L_n be the length of the interval of the u-axis over which it lies. If S_n lies on or below $v = 1$, at least one point of S_n with given abscissa can be joined by a vertical curve to a boundary of $\mathfrak{G}_n$ (which may be S_n) below $v = 1$. Lemma 3 yields

$$L_n \leqq \frac{2D(\mathfrak{x}_n)}{t_n^2}$$

where t_n is the minimum oscillation of $\mathfrak{x}_n$ on such vertical curves. By Lemma 5, t_n cannot $\to 0$, so that L_n cannot $\to \infty$.

If S_n has the height $v_n > 1$, all the infinite slits lie below S_n and each

point of the lower edge of S_n can be joined by a vertical curve in $\mathfrak{G}_n$ to another boundary of $\mathfrak{G}_n$. The oscillation of $\mathfrak{x}_n$ on each such vertical curve is $\geqq \delta$, and Lemma 3 yields

$$\frac{L_n}{v_n} \leqq \frac{2D(\mathfrak{x}_n)}{\delta^2}.$$

Hence if $L_n \to \infty$, v_n must $\to \infty$. Thus, the length of a boundary can become infinite only if its height tends to infinity.

Suppose that some boundary slits, or vertices of infinite slits, converged to the point at infinity. Enclose all the boundary slits, and all the vertices, which remain in a bounded portion of the plane by a fixed square R of side $2M$, $M > 1$, with the origin as center; and let R_n be a concentric square of side $2M_n \to \infty$ such that no boundaries or vertices occur between R and R_n. Consider those parts of concentric squares between R and R_n which join boundary points on the u-axis, and let W_n with end points A_n, B_n be that square on which the oscillation of $\mathfrak{x}_n$ is its minimum t_n. By Lemma 2,

$$\frac{t_n^2}{64} \log \frac{M_n}{M} \leq D(\mathfrak{x}_n),$$

which implies that $t_n \to 0$, and in particular that $|\mathfrak{x}_n(A_n) - \mathfrak{x}_n(B_n)| \to 0$. Hence the oscillation of $\mathfrak{x}_n$ on one of the arcs $\overparen{A_n 0 B_n}$, $\overparen{A_n \infty B_n}$ of the boundary on the u-axis tends to zero. The oscillation of $\mathfrak{x}_n$ on one of the closed curves $W_n \overparen{A_n 0 B_n}$, $W_n \overparen{A_n \infty B_n}$ therefore tends to zero. Now, by the normalization of $\mathfrak{G}_n$, $W_n \overparen{A_n 0 B_n}$ contains in its interior either boundary slits or vertices of infinite slits. If it contains a boundary slit, $W_n \overparen{A_n 0 B_n}$ is non-bounding. If it contains no boundary slits, there is a non-bounding closed curve C_n in its interior; the oscillation of $\mathfrak{x}_n$ on C_n is less than the oscillation of $\mathfrak{x}_n$ on $W_n \overparen{A_n 0 B_n}$ by the maximum-minimum principle of potential theory. Similarly for $W_n \overparen{A_n \infty B_n}$ since there are boundaries or vertices inside it. We would therefore have non-bounding closed curves on which the oscillation of $\mathfrak{x}_n$ tends to zero, which is a contradiction. Neither boundaries nor vertices of infinite slits can tend to infinity.

Thus, the domains $\mathfrak{G}_n$ converge to a slit domain $\mathfrak{G}$.

(b) *The domain $\mathfrak{G}$ has the prescribed topological structure.* Construct the domain $\bar{\mathfrak{G}}_n$ as in section 3. In this construction, there are no remaining strips. Otherwise, by attaching remaining strips together along coördinated horizontal edges to form a cycle, the oscillation of $\mathfrak{x}_n$ on at least one vertical curve in this cycle would tend to zero, by Lemma 3, contradicting Lemma 5.

About any vertex of $\bar{\mathfrak{G}}_n$, e. g., where there are small irregular edges, draw a complete neighborhood, containing the small irregular edges, bounded by a curve consisting of pieces of squares of side ϵ_n, $\epsilon_n \to 0$ as $n \to \infty$. Consider all curves consisting of concentric squares of sides between ϵ_n and ϵ, ϵ fixed, and let l_n be the minimum oscillation of $\mathfrak{x}_n$ on these curves. Here, ϵ_n and ϵ are determined so that between the corresponding curves the domain $\bar{\mathfrak{G}}_n$ is homogeneous and of the same structure as $\mathfrak{G}$. If m is the number of pieces of squares in each such curve, Lemma 2 yields

$$\frac{l_n{}^2}{64m^2} \log \frac{\epsilon}{\epsilon_n} \leqq D(\mathfrak{x}_n)$$

so that $l_n \to 0$ as $n \to \infty$. The curves therefore cannot join distinct boundaries of $\mathfrak{G}_n$; if they are closed, they must bound in $\mathfrak{G}_n$; if they join points of one boundary, they must form together with arcs of this boundary closed curves which bound in $\mathfrak{G}_n$. It follows, as in Lemma 1, that $\mathfrak{G}$ has the prescribed topological structure.

8. Convergence of the minimizing sequence. Let $S_1, S_2, \cdots, S_n, \cdots$ be corresponding boundaries of $\mathfrak{G}_1, \mathfrak{G}_2, \cdots, \mathfrak{G}_n, \cdots$ respectively, tending to the boundary S of $\mathfrak{G}$. We shall show that: *the boundary values of* $\mathfrak{x}_1, \mathfrak{x}_2, \cdots, \mathfrak{x}_n, \cdots$ *on* $S_1, S_2, \cdots, S_n, \cdots$ *respectively are equicontinuous.* If not, there is a subsequence which we rename $\mathfrak{x}_1, \mathfrak{x}_2, \cdots, \mathfrak{x}_n, \cdots$ a positive constant γ, and points P_n, Q_n on S_n for which $\overline{P_nQ_n} \to 0$ as $n \to \infty$, such that $|\mathfrak{x}_n(P_n) - \mathfrak{x}_n(Q_n)| > \gamma$.[21] By choosing a subsequence again if necessary, we may suppose that $P_n \to R$, $Q_n \to R$. Construct the domain $\bar{\mathfrak{G}}_n$, and about P_n in $\bar{\mathfrak{G}}_n$ construct a curve consisting of squares of side ϵ_n, where $\epsilon_n \to 0$ as $n \to \infty$, containing Q_n in its interior. Lemma 2 shows that the minimum oscillation l_n of $\mathfrak{x}_n$ on curves consisting of concentric squares of sides between ϵ_n and ϵ, ϵ fixed, tends to zero as $n \to \infty$. Let W_n be that curve on which $\mathfrak{x}_n$ has the oscillation l_n, and let A_n, B_n be the end points of W_n on S_n. Since $l_n \to 0$, $|\mathfrak{x}_n(A_n) - \mathfrak{x}_n(B_n)| \to 0$, so that the oscillation of $\mathfrak{x}_n$ on one of the two arcs bounded by A_n, B_n tends to zero. On the small arc $\widehat{A_nP_nQ_nB_n}$ the oscillation of $\mathfrak{x}_n$ is $> \gamma$; therefore the oscillation of $\mathfrak{x}_n$ on the large arc $\widehat{A_nB_n}$ tends to zero, and so does the oscillation of $\mathfrak{x}_n$ on $W_n\widehat{A_nB_n}$. But the closed curve $W_n\widehat{A_nB_n}$ is non-bounding in $\mathfrak{G}_n$. This is a contradiction, and our statement is proved.

The boundary values of $\mathfrak{x}_n$ on the u-axis are also equicontinuous. For, if

---

[21] A set of functions $f_n(P)$, $n = 1, 2, \cdots$, is equicontinuous if, for any ϵ, there is a δ such that $\overline{PQ} < \delta$ implies $|f_n(P) - f_n(Q)| < \epsilon$ for every n. The non-equicontinuity of a set $f_n(P)$ is easily put in the form stated in the text.

$|\mathfrak{x}_n(P_n) - \mathfrak{x}_n(Q_n)| > \gamma$ where $P_n \to \infty$, $Q_n \to \infty$ (the case $P_n, Q_n \to R \neq \infty$ is identical with the above), construct a square with origin as center of side $M_n \to \infty$, not containing P_n, Q_n in its interior. Consider concentric squares of sides between M_n and M, M fixed and containing all the boundaries and branch points in its interior, and let W_n with end points A_n, B_n be that square on which $\mathfrak{x}_n$ has its minimum oscillation. Exactly as in the previous case, the oscillation of $\mathfrak{x}_n$ on $W_n\widehat{A_n 0 B_n}$ tends to zero and $W_n\widehat{A_n 0 B_n}$ is non-bounding, which is a contradiction.

The boundary values are thus equicontinuous, and a subsequence which we rename $\mathfrak{x}_1, \mathfrak{x}_2, \cdots, \mathfrak{x}_n, \cdots$ can be found whose boundary values converge uniformly. We now show that: *the potential surfaces* $\mathfrak{x}_1, \mathfrak{x}_2, \cdots, \mathfrak{x}_n, \cdots$ *in the domains* $\mathfrak{G}_1, \mathfrak{G}_2, \cdots, \mathfrak{G}_n, \cdots$ *respectively, converge uniformly to a potential surface* $\mathfrak{x}(u, v)$ *in the domain* $\mathfrak{G}$ *which maps the boundaries of* $\mathfrak{G}$ *onto the curves* $\Gamma_1, \cdots, \Gamma_k$. Construct the domain $\bar{\mathfrak{G}}_n$ as in section 3. Take duplicate copies of $\bar{\mathfrak{G}}_n$ and adjoin them to $\bar{\mathfrak{G}}_n$ along a pair of coördinated horizontal edges, for each edge of $\bar{\mathfrak{G}}_n$, and denote the resulting domain covering $\bar{\mathfrak{G}}_n$ several times by $\bar{\bar{\mathfrak{G}}}_n$. As $n \to \infty$, $\bar{\bar{\mathfrak{G}}}_n$ converges to a domain $\bar{\bar{\mathfrak{G}}}$ which covers $\mathfrak{G}$ several times. The surface $\mathfrak{x}_n$ is potential on $\bar{\mathfrak{G}}_n$ and also on $\bar{\bar{\mathfrak{G}}}_n$: it converges uniformly to a potential surface $\mathfrak{x}(u, v)$ in $\bar{\bar{\mathfrak{G}}}$ (the small irregular edges of $\bar{\bar{\mathfrak{G}}}_n$ occurring at branch points may be surrounded by squares of radius η_n, $\eta_n \to 0$, on which the oscillation of $\mathfrak{x}_n$ tends to zero). But $\bar{\bar{\mathfrak{G}}}$ contains $\mathfrak{G}$, so that $\mathfrak{x}(u, v)$ is a potential surface in $\mathfrak{G}$ mapping the boundaries of $\mathfrak{G}$ onto the curves $\Gamma_1, \cdots, \Gamma_k$.[22]

The derivatives of $\mathfrak{x}_n$ converge uniformly in any interior closed subdomain H of $\mathfrak{G}$ to the derivatives of $\mathfrak{x}$, so that

$$D_H(\mathfrak{x}) = \lim_{n\to\infty} D_H(\mathfrak{x}_n) \leqq \lim_{n\to\infty} D_{\mathfrak{G}_n}(\mathfrak{x}_n) = d_{\alpha,\beta}.$$

Letting H tend to $\mathfrak{G}$, we have[23] $D_{\mathfrak{G}}(\mathfrak{x}) \leqq d_{\alpha,\beta}$. But $\mathfrak{x}(u, v)$ is an allowable

[22] The developments thus far also yield a proof of the following. Let $\mathfrak{G}_1, \mathfrak{G}_2, \cdots, \mathfrak{G}_n, \cdots$ be a sequence of slit domains of given topological structure tending to a slit domain $\mathfrak{G}$ of the same topological structure. Suppose that each $\mathfrak{G}_n$ is mapped conformally into a slit domain $\mathfrak{G}'_n$ by an analytic function $f_n(z)$. Then a subsequence $\mathfrak{G}'_{n_i}$ tends to a slit domain $\mathfrak{G}'$ of the same topological structure, and $f_{n_i}(z)$ converges uniformly to an analytic function $f(z)$ which maps $\mathfrak{G}$ into $\mathfrak{G}'$.

This permits defining convergence of Riemann surfaces. The Riemann surfaces $\mathfrak{R}_1, \mathfrak{R}_2, \cdots, \mathfrak{R}_n, \cdots$ converge to the Riemann surface $\mathfrak{R}$ if slit domains $\mathfrak{G}_1, \mathfrak{G}_2, \cdots, \mathfrak{G}_n, \cdots$ conformally equivalent to $\mathfrak{R}_1, \mathfrak{R}_2, \cdots, \mathfrak{R}_n, \cdots$ respectively has all its limit slit domains $\mathfrak{G}$ conformally equivalent to $\mathfrak{R}$.

[23] This states the lower semi-continuity of the Dirichlet functional in the class of potential surfaces.

surface whose inner diameter, being the limit of the inner diameters of $\mathfrak{x}_n$, lies between α and β. Therefore $D_{\mathfrak{G}}(\mathfrak{x}) \geqq d_{\alpha,\beta}$ and finally $D(\mathfrak{x}) = d_{\alpha,\beta}$. The potential surface $\mathfrak{x}(u,v)$ is the solution to our minimum problem.

THE SOLUTION IS A MINIMAL SURFACE.

9. The variational condition. We shall first suppose that the slits in the domain $\mathfrak{G}$ over which the surface $\mathfrak{x}(u,v)$ is defined are completely separated. The general case will be considered in section 13.

Let $\lambda(u,v)$, $\mu(u,v)$ be any two functions, continuous and having bounded piecewise continuous first derivatives,

$$|\lambda_u|,\ |\lambda_v|,\ |\mu_u|,\ |\mu_v| < M.$$

Suppose that $\lambda(u,v)$ takes equal values on coördinated points of $\mathfrak{G}$, and that $\mu(u,v)$ takes constant values on each slit of $\mathfrak{G}$ (not necessarily equal values on coördinated slits). Then the transformation

$$(T)\qquad \begin{cases} U = u + \epsilon\lambda(u,v) \\ V = v + \epsilon\mu(u,v), \qquad |\epsilon| < \dfrac{1}{4M}, \end{cases}$$

maps $\mathfrak{G}$ in a one-to-one manner into a slit domain $\mathfrak{G}_\epsilon$ of the (U,V)-plane. For, the Jacobian of the transformation is

$$J = 1 + \epsilon(\lambda_u + \mu_v) + \epsilon^2(\lambda_u\mu_v - \lambda_v\mu_u) > \tfrac{1}{4},$$

and slits go into slits, coördinated points into coördinated points. The surface $\mathfrak{z}_\epsilon(U,V)$ defined over the slit domain $\mathfrak{G}_\epsilon$ by

$$\mathfrak{z}_\epsilon(U,V) = \mathfrak{x}(u,v)$$

is an allowable surface whose diameter, being equal to that of $\mathfrak{x}(u,v)$, lies between α and β. It is therefore necessary that

$$D_{\mathfrak{G}_\epsilon}(\mathfrak{z}_\epsilon) \geqq D_{\mathfrak{G}}(\mathfrak{x}).$$

A simple calculation yields

$$D_{\mathfrak{G}_\epsilon}(\mathfrak{z}_\epsilon) = \tfrac{1}{2}\iint\limits_{\mathfrak{G}_\epsilon}\left[\left(\frac{\partial\mathfrak{z}_\epsilon}{\partial U}\right)^2 + \left(\frac{\partial\mathfrak{z}_\epsilon}{\partial V}\right)^2\right]dU\,dV$$

$$= \tfrac{1}{2}\iint\limits_{\mathfrak{G}}\{[\mathfrak{x}_u(1+\epsilon\mu_v) - \epsilon\mu_u\mathfrak{x}_v]^2 + [-\epsilon\lambda_v\mathfrak{x}_u + \mathfrak{x}_v(1+\epsilon\lambda_u)]^2\}\frac{du\,dv}{J}$$

$$D_{\mathfrak{G}_\epsilon}(\mathfrak{z}_\epsilon) = D_{\mathfrak{G}}(\mathfrak{x}) + \frac{\epsilon}{2}\iint\limits_{\mathfrak{G}}[\mu_v(\mathfrak{x}_u^2 - \mathfrak{x}_v^2) - \mu_u\cdot 2\mathfrak{x}_u\mathfrak{x}_v]\,du\,dv$$

$$-\frac{\epsilon}{2}\iint\limits_{\mathfrak{G}}[\lambda_u(\mathfrak{x}_u^2 - \mathfrak{x}_v^2) + \lambda_v\cdot 2\mathfrak{x}_u\mathfrak{x}_v]\,du\,dv + \epsilon^2 I$$

where I is an integral which is uniformly bounded for all ϵ, $|I| < 16M^2 D_{\mathfrak{G}}(\mathfrak{x})$. Exactly,

$$I = \tfrac{1}{2} \iint\limits_{\mathfrak{G}} [(\lambda_u^2 + \lambda_v^2)\mathfrak{x}_u^2 + (\lambda_u\mu_u + \lambda_v\mu_v) \cdot 2\mathfrak{x}_u\mathfrak{x}_v + (\mu_u^2 + \mu_v^2)\mathfrak{x}_v^2] \frac{dudv}{J}$$

$$+ \tfrac{1}{2} \iint\limits_{\mathfrak{G}} [\mathfrak{x}_u^2 + \mathfrak{x}_v^2 + \epsilon(\mu_v - \lambda_u)(\mathfrak{x}_u^2 - \mathfrak{x}_v^2)$$

$$- \epsilon(\lambda_v + \mu_u) \cdot 2\mathfrak{x}_u\mathfrak{x}_v] \frac{\lambda_v\mu_u - \lambda_u\mu_v}{J} dudv.$$

Hence, by the usual argument, we must have

$$\iint\limits_{\mathfrak{G}} [\mu_v(\mathfrak{x}_u^2 - \mathfrak{x}_v^2) - \mu_u \cdot 2\mathfrak{x}_u\mathfrak{x}_v]dudv$$

$$- \iint\limits_{\mathfrak{G}} [\lambda_u(\mathfrak{x}_u^2 - \mathfrak{x}_v^2) + \lambda_v \cdot 2\mathfrak{x}_u\mathfrak{x}_v]dudv = 0.$$

Approximate $\mathfrak{G}$, cut along its slits, by a sequence of closed subdomains $\mathfrak{G}_\nu$ of $\mathfrak{G}$ with smooth boundaries C_ν. By integration by parts, the above condition becomes

$$\int\limits_{C_\nu} \mu[(\mathfrak{x}_u^2 - \mathfrak{x}_v^2)du + 2\mathfrak{x}_u\mathfrak{x}_v dv] - \int\limits_{C_\nu} \lambda[2\mathfrak{x}_u\mathfrak{x}_v du - (\mathfrak{x}_u^2 - \mathfrak{x}_v^2)dv] \to 0$$

as $\mathfrak{G}_\nu \to \mathfrak{G}$, where the direction of integration along the boundaries C_ν is such that $\mathfrak{G}_\nu$ lies to the left of C_ν. The additional double integral term vanishes since $\mathfrak{x}(u, v)$ is a potential surface.

Set

$$\phi(w) = (\mathfrak{x}_u - i\mathfrak{x}_v)^2 = \mathfrak{x}_u^2 - \mathfrak{x}_v^2 - 2i\mathfrak{x}_u\mathfrak{x}_v,$$

where $w = u + iv$; $\phi(w)$ is a single-valued analytic function of w in the domain $\mathfrak{G}$. It takes equal values on coördinated infinite slits of the first kind, and conjugate complex values on coördinated infinite slits of the second kind. The condition above can be written

$$(V) \qquad \int\limits_{C_\nu} \{\mu \mathfrak{R}(\phi(w)dw) + \lambda \mathfrak{I}(\phi(w)dw)\} \to 0$$

as $\mathfrak{G}_\nu \to \mathfrak{G}$, where $\mathfrak{R}$ and $\mathfrak{I}$ denote the real and imaginary components of a complex number. This is the desired variational condition, from which we shall conclude that $\phi(w) \equiv 0$.

10. Special variations. We shall vary the boundary values of $\mathfrak{x}(u, v)$, the position of the boundary slits, and the position of the infinite slits individually by choosing particular functions $\lambda(u, v)$, $\mu(u, v)$.[24]

---

[24] Since the normalization of the slit domains are of an unessential type, no special care need be taken on its account.

(a) *Boundary values.* Let S be a boundary slit of $\mathfrak{G}$ at $v = v_0, u_1 \leqq u \leqq u_2$. Set $\mu(u, v) \equiv 0$, and $\lambda(u, v) = 0$ outside the rectangle $u_1 \leqq u \leqq u_2$, $v_0 \leqq v \leqq v_0 + \eta$ but arbitrary inside. The variational condition (V) is

$$\int_{\substack{u_1 \\ v=v_0+\epsilon}}^{u_2} \lambda \mathfrak{I}(\phi(w))du \to 0 \text{ as } \epsilon \to 0.$$

By Lemma 4, this implies that $\mathfrak{I}(\phi(w)) = 0$ on the top edge of S. Similarly for the bottom edge.

For the end-point $w_1 = u_1 + iv_0$ of S, choose $\mu(u, v)$ and $\lambda(u, v)$ both $= 0$ outside the circle $|w - w_1| = \eta^2$. Inside the circle make the transformation $\sqrt{w - w_1} = \bar{w} = \bar{u} + i\bar{v}$ which carries the circle into a semicircle of radius η of the $\bar{w}$-plane, the line $\bar{v} = 0$ corresponding to the boundary slit. Choose $\bar{\mu}(\bar{u}, \bar{v}) \equiv 0$, and $\bar{\lambda}(\bar{u}, \bar{v}) = 0$ outside the rectangle $-\eta' \leq \bar{u} \leq \eta'$, $0 \leq \bar{v} \leq \delta$ but arbitrary inside, where $\eta' < \eta$ and the rectangle lies entirely inside the semicircle. The variational condition (V) is

$$\int_{\substack{-\eta' \\ \bar{v}=\epsilon}}^{\eta'} \bar{\lambda} \mathfrak{I}(\bar{\phi}(\bar{w}))d\bar{u} \to 0 \text{ as } \epsilon \to 0,$$

where

$$\bar{\phi}(\bar{w}) = (\mathfrak{x}_{\bar{u}} - i\mathfrak{x}_{\bar{v}})^2 = (\mathfrak{x}_u - i\mathfrak{x}_v)^2 \left(\frac{dw}{d\bar{w}}\right)^2 = 4(w - w_1)\phi(w).$$

Lemma 4 yields $\mathfrak{I}(\bar{\phi}(\bar{w})) = 0$ on $\bar{v} = 0$, or

$$\mathfrak{I}[(w - w_1)\phi(w)] \to 0 \text{ as } w \to w_1.$$

A similar reasoning applies for the end-point $w_2 = u_2 + iv_0$. Thus,[25]

$$(1) \quad \begin{cases} \mathfrak{I}(\phi(w)) = 0 \text{ for each edge of a boundary} \\ \mathfrak{I}[(w - w_0)\phi(w)] \to 0 \text{ as } w \to w_0 \text{ for each end-point } w_0 \text{ of a boundary slit} \end{cases}$$

(b) *Boundary slits.* The height of the boundary slit S and the position of its end-points w_1, w_2 must be varied. Choose 1) $\lambda(u, v) \equiv 0$, $\mu(u, v) = 1$ near S and $\mu(u, v) = 0$ outside a neighborhood of S; 2) $\mu(u, v) \equiv 0$, $\lambda(u, v) = 1$ near w_1 and $\lambda(u, v) = 0$ outside a neighborhood of w_1; 3) the same as 2) for w_2. Equation (V) becomes

[25] The relations (1) can be obtained in one step by making the transformation

$$w - w_0 = \frac{l}{4}\left(\bar{w} + \frac{1}{\bar{w}}\right),$$

where w_0 is the center of the slit S and l its length, which maps the slit domain into the unit circle.

$$(2)\qquad \begin{cases} \mathfrak{R} \displaystyle\int_C \phi(w)\,dw = 0 \\ \mathfrak{I} \displaystyle\int_{C_1^{(\epsilon)}} \phi(w)\,dw \to 0 \text{ as } \epsilon \to 0 \\ \mathfrak{I} \displaystyle\int_{C_2^{(\epsilon)}} \phi(w)\,dw \to 0 \text{ as } \epsilon \to 0 \end{cases}$$

respectively, where C is a curve surrounding S, and $C_1^{(\epsilon)}, C_2^{(\epsilon)}$ are circles of radius ϵ surrounding w_1, w_2.

(c) *Infinite slits.* Let S_1, S_2 be a pair of coördinated infinite slits, and C_1, C_2 curves surrounding them. The height of S_1, the height of S_2, and the abscissa of their end-points must be varied. Set 1) $\lambda(u, v) \equiv 0$, $\mu(u, v) = 1$ near S_1 and $\mu(u, v) = 0$ outside a neighborhood of S_1; 2) same as 1) for S_2; 3) $\mu(u, v) \equiv 0$, $\lambda(u, v) = 1$ near S_1, S_2 and $\lambda(u. v) = 0$ outside a neighborhood of them. Equation (V) yields

$$(3)\qquad \begin{cases} \mathfrak{R} \displaystyle\int_{C_1} \phi(w)\,dw = 0 \\ \mathfrak{R} \displaystyle\int_{C_2} \phi(w)\,dw = 0 \\ \mathfrak{I} \displaystyle\int_{C_1} \phi(w)\,dw + \mathfrak{I} \displaystyle\int_{C_2} \phi(w)\,dw = 0. \end{cases}$$

11. Other forms of the variational conditions.

(a) *Boundary values.* Let S be the boundary slit with end-points w_1, w_2. The first of the conditions (1) shows that $\phi(w)$ can be extended analytically across S to a double sheeted Riemann surface $\mathfrak{S}^*$ with branch points at w_1 and w_2. In the neighborhood of w_1, w_2 the analytic function $\phi(w)$ can be expanded into power series in $\sqrt{w - w_1}$, $\sqrt{w - w_2}$. The second of the conditions (1) shows that the power series are

$$(1')\qquad \begin{cases} \phi(w) = \dfrac{a_1}{w - w_1} + \dfrac{b_1}{\sqrt{w - w_1}} + c_1 + d_1\sqrt{w - w_1} + \cdots \text{ about } w_1 \\ \phi(w) = \dfrac{a_2}{w - w_2} + \dfrac{b_2}{\sqrt{w - w_2}} + c_2 + d_2\sqrt{w - w_2} + \cdots \text{ about } w_2 \\ \qquad a_1, a_2, b_1, b_2, \cdots \text{ real numbers.} \end{cases}$$

(b) *Boundary slits.* The conditions (2) become

$$(2')\qquad \begin{cases} \mathfrak{R} \displaystyle\int_C \phi(w)\,dw = 0 \\ a_1 = 0 \\ a_2 = 0. \end{cases}$$

(c) *Infinite slits.* Let S_1, S_2 be a pair of coördinated infinite slits with end-points at W_1, W_2. The potential surface $\mathfrak{x}(u, v)$ is also a regular potential surface on the double sheeted surface obtained by attaching a duplicate of $\mathfrak{G}$ along the edges of S_2 to $\mathfrak{G}$ along the corresponding edges of S_1. It follows that

$$\phi(w) = \frac{A}{w - W_1} + \frac{B}{\sqrt{w - W_1}} + C + D\sqrt{w - W_1} + \cdots \quad \text{about } W_1$$

$$\left.\begin{aligned} \text{and } \phi(w) &= \frac{A}{w - W_2} - \frac{B}{\sqrt{w - W_2}} + C - D\sqrt{w - W_2} + \cdots \\ \text{or } \phi(w) &= \frac{\bar{A}}{w - W_2} + \frac{\bar{B}}{\sqrt{w - W_2}} + \bar{C} + \bar{D}\sqrt{w - W_2} + \cdots \end{aligned}\right\} \text{about } W_2$$

where the bar denotes the conjugate complex, according as the slits are of the first kind or second kind respectively.

The curves C_1, C_2 may be collapsed into curves which run along the edges of the slits till $u = U_0 + \epsilon$ and around circles of radius ϵ with W_1, W_2 as centers. Taking into account the coördinations of the edges $S_1^+, S_1^-, S_2^+, S_2^-$ of the slits S_1, S_2 and the values of $\phi(w)$ on corresponding edges, and letting $\epsilon \to 0$, the equations (3) become

$$(3') \qquad \begin{cases} \mathfrak{R} \displaystyle\int_{C_1} \phi(w)\,dw = \mathfrak{R} \int_{C_2} \phi(w)\,dw = 0 \\ \mathfrak{I}(A) = 0 \\ \mathfrak{R}(A) = 0. \end{cases}$$

12. Proof that $\phi(w) \equiv 0$. The equations (3′) and the first of (2′) show that $\mathfrak{R}\int_C \phi(w)\,dw = 0$ about any closed curve C in $\mathfrak{G}$. For, this holds if C winds around a boundary slit, or around an infinite slit, or around a branch point of an infinite slit; and the integral around any closed curve is a linear combination of these. Furthermore, $\mathfrak{R}\int_C \phi(w)\,dw = 0$ for any closed curve C in the double sheeted Riemann surface $\mathfrak{G}^*$ with a boundary slit S as branch line and its end-points w_1, w_2 as branch points. For it holds, by (1′) and the last two equations of (2′), if C winds around the branch point w_1 or w_2. Thus, the potential function $\mathfrak{R}\int_{w_0}^{w} \phi(w)\,dw$ where w_0 is any point, is single valued in $\mathfrak{G}$ and in any $\mathfrak{G}^*$.

The potential function $\mathfrak{R}\int_{w_0}^{w} \phi(w)\,dw$ has no singularities in $\mathfrak{G}$. For it is regular at all end-points of slits by virtue of (2′), (3′) $(A = a_1 = a_2 = 0)$; it is regular at infinity as the following argument shows. Let R be the largest modulus of the vertices of the slit domain, so that the region outside

the circle $|w| = R$ contains no vertices. Since $\mathfrak{I}(\phi(w)) = 0$ on the u-axis, $\phi(w)$ is defined on the whole (u, v)-plane by reflection. By the mean value property,

$$\phi(w) = \frac{1}{\pi(|w| - R)^2} \int\int \phi(w) r dr d\theta$$

the integral being taken over a circle of radius $|w| - R$ and center w, so that

$$|\phi(w)| \leqq \frac{1}{\pi(|w| - R)^2} \int\int |\phi(w)| r dr d\theta \leqq \frac{2D(\mathfrak{x})}{\pi(|w| - R)^2}.$$

Hence $\mathfrak{R} \int_{w_0}^{w} \phi(w) dw$ is bounded and regular at infinity.

The potential function $\mathfrak{R} \int_{w_0}^{w} \phi(w) dw$, having no singularities in $\mathfrak{G}$, must take its maximum (and minimum) on a boundary of $\mathfrak{G}$. But, $\mathfrak{I} \int_{w_0}^{w} \phi(w) dw$ has constant values on each boundary of $\mathfrak{G}$. For, the difference of $\mathfrak{I} \int_{w_0}^{w} \phi(w) dw$ for two points of a boundary is the integral of $\mathfrak{I}(\phi(w))$ along the boundary, which vanishes by virtue of (1). Hence $\mathfrak{R} \int_{w_0}^{w} \phi(w) dw$ takes equal values on reflected points in $\mathfrak{G}^*$. It follows that the maximum (or minimum) of $\mathfrak{R} \int_{w_0}^{w} \phi(w) dw$ cannot be on a boundary of $\mathfrak{G}$ unless

$$\mathfrak{R} \int_{w_0}^{w} \phi(w) dw = \text{constant.}^{26}$$

This implies that $\phi(w)$ vanishes identically.

Thus, $E - G = 0$, $F = 0$, and the surface $\mathfrak{x} = \mathfrak{x}(u, v)$ is a minimal surface. Q. E. D.

13. The general domain. It remains to consider the case when the slits of the domain $\mathfrak{G}$ over which $\mathfrak{x}(u, v)$ is defined are not separated. It is then necessary to perform variations which separate the slits. Suppose first that only infinite slits have coalesced, the boundary slits being distinct. Then the equations (1), (2) and (1′), (2′) are valid, but (3), (3′) must be modified. Let W be a branch point of $\mathfrak{G}$ of order ν with a vertices $P_1, \cdots, P_a$ of angle 2π and $2c$ vertices $Q_1, Q'_1, \cdots, Q_c, Q'_c$ of angle π, where the notation is the same as in section 2, page 860. Let S_j be the slit with P_j as vertex, and E_i, F_i

[26] The conditions on $\mathfrak{R} \int_{w_0}^{w} \phi(w) dw$ may be interpreted to mean that it is regular and single valued on the double of $\mathfrak{G}$. Then the equation $\mathfrak{R} \int_{w_0}^{w} \phi(w) dw = \text{constant}$ is also immediate.

and E'_i, F_i the edges with Q_i and Q'_i as vertices; and choose the $\nu+1$ curves C_j and K_i as in section 2, page 860. The variational conditions are

$$(4)\quad \begin{cases} \mathfrak{R}\displaystyle\int_{C_j}\phi(w)\,dw = 0, & (j=1,2,\cdots,a) \\ \mathfrak{R}\displaystyle\int_{K_i}\phi(w)\,dw = 0, & (i=1,2,\cdots,c) \\ \mathfrak{I}\displaystyle\int_{N}\phi(w)\,dw = 0 & \\ \displaystyle\int_{N}\kappa(w)\phi(w)\,dw = 0,\ \kappa(w) = (w-W)^{\frac{\nu-1}{\nu+1}}, (w-W)^{\frac{\nu-2}{\nu+1}}, \cdots, (w-W)^{\frac{1}{\nu+1}} & \end{cases}$$

where N is a curve completely encircling the point W in $\mathfrak{S}$.[27]

These are obtained from (V) by selecting $\lambda(u,v)$, $\mu(u,v)$ as follows: 1) $\lambda(u,v)\equiv 0$, $\mu(u,v)=1$ near S_j and $\mu(u,v)=0$ outside a neighborhood of S_j; 2) $\lambda(u,v)\equiv 0$, $\mu(u,v)=1$ near E_i, E'_i, $\mu(u,v)=0$ outside a neighborhood of them, and taking equal values on coördinated points of F_i; 3) $\mu(u,v)\equiv 0$, $\lambda(u,v)=1$ near all the points $P_1,\cdots,P_a, Q_1, Q'_1,\cdots,Q_c, Q'_c$ and $\lambda(u,v)=0$ outside a neighborhood of them; 4) $\lambda(u,v)=\mathfrak{R}(\kappa(w))$ and $\mu(u,v)=\mathfrak{I}(\kappa(w))$ near W, and both $=0$ outside a neighborhood of W, where $\kappa(w)$ is any function analytic in $\mathfrak{S}$ in the neighborhood of $w=W$ and vanishing for $w=W$. In the choice (2), the domain $\mathfrak{S}_\epsilon$ is not a slit domain but it may be pieced together to form a slit domain, as Figure 4 below illustrates. In the choice 4), the domain $\mathfrak{S}_\epsilon$ may also be pieced together to form a slit domain; furthermore, near $w=W$, $\lambda_u=\mu_v$, $\lambda_v=-\mu_u$, $J=(1+\epsilon\lambda_u)^2+\epsilon^2\lambda_v^2\geqq 0$, and the integrand of I vanishes. The condition (V) is therefore valid. This choice 4) yields $\mathfrak{I}\int_N \kappa(w)\phi(w)\,dw=0$; replacing $\kappa(w)$ by $i\kappa(w)$ the result is $\mathfrak{R}\int_N \kappa(w)\phi(w)\,dw=0$, and finally, $\int_N \kappa(w)\phi(w)\,dw=0$, which will be used only for the functions $\kappa(w)$ listed in (4).

---

[27] Suppose that the slit domain has no boundaries and is of genus p, $k=0$, $q=2p-2$. Then $\phi(w)$ is a single valued function on a Riemann surface of genus p with a pole of order 2ν at each branch point of order ν; $\phi(w)$ is therefore representable linearly in terms of the elementary functions on the Riemann surface with $\Sigma\, 4\nu = 4\rho = 8p$ arbitrary constants. The equations (4) are 3ν conditions, or $\Sigma\, 3\nu = 3\rho = 6p$ conditions for all branch points; $2p$ additional conditions are obtained from the fact that $\phi(w)$ is single valued. We have, therefore, a system of $8p$ linear homogeneous equations in the $8p$ constants, from which it can be concluded 'in general' that the constants are all zero, or $\phi(w)\equiv 0$. In the general case with boundaries, it can be seen by noting that $\phi(w)$ is single valued on the double of the slit domain, that the number of equations is equal to the number of constants.

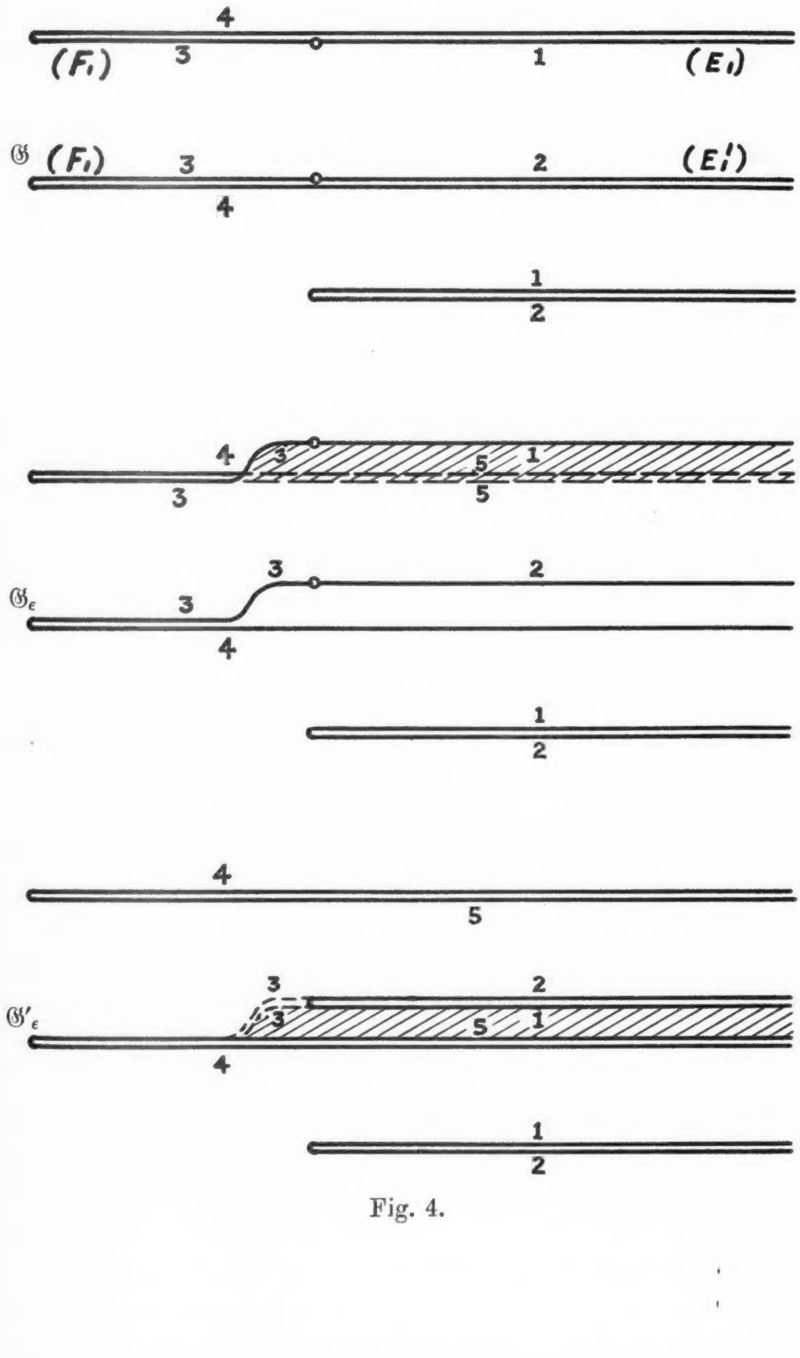


Fig. 4.

Now, the curves C_j, K_i can be collapsed into curves running around the points $P_1, \cdots, P_a, Q_1, Q'_1, \cdots, Q_c, Q'_c$ along the edges $S_1^+, S_1^-, \cdots, S_a^+, S_a^-$, $E_1, E'_1, \cdots, E_c, E'_c$, and along curves $\bar{K}_i$ which surround branch points different from $w = W$. By suitably adding or subtracting the first $a + c$ equations of (4) (adding in case all the relevent edges are of the first kind), the integrals along the edges $S_1^+, S_1^-, \cdots, E_c, E'_c$ cancel each other, while the integrals along $\bar{K}_i$ vanish by virtue of the same relations (4) for all other branch points. There remains $\mathfrak{R} \int_N \phi(w)\,dw = 0$. This, together with the third of the equations (4), shows that the last equation of (4) holds also for $\kappa(w) = 1$.

The potential function $\mathfrak{x}(u, v)$ is regular in the neighborhood of $w = W$, and is the real part of a power series in $(w - W)^{\frac{1}{\nu+1}}$. Therefore,

$$\phi(w) = \frac{A_{2\nu}}{(w-W)^{\frac{2\nu}{\nu+1}}} + \frac{A_{2\nu-1}}{(w-W)^{\frac{2\nu-1}{\nu+1}}} + \cdots + \frac{A_\nu}{(w-W)^{\frac{\nu}{\nu+1}}} + \cdots + \frac{A_1}{(w-W)^{\frac{1}{\nu+1}}} + A_0 + \cdots$$

about $w = W$. The last of the conditions (4) for the functions

$$\kappa(w) = (w-W)^{\frac{\nu-1}{\nu+1}},\ (w-W)^{\frac{\nu-2}{\nu+1}}, \cdots, (w-W)^{\frac{1}{\nu+1}},\ 1$$

yields successively $A_{2\nu} = 0$, $A_{2\nu-1} = 0, \cdots, A_{\nu+2} = 0$, $A_{\nu+1} = 0$. This shows that $\mathfrak{R} \int_{w_0}^{w} \phi(w)\,dw$ has no singularities in $\mathfrak{G}$. The first two relations of (4), and (2′), show that $\mathfrak{R} \int_{w_0}^{w} \phi(w)\,dw$ is single valued. It follows, as in section 12, that $\phi(w) \equiv 0$.

Consider, now, the most general case in which boundary slits have coalesced with infinite slits. Let S be such a boundary slit, and let $w = \Omega$ be a branch point of order μ on S, i. e., a complete neighborhood of it has an angle $2(\mu + \frac{1}{2})\pi$. Let C_j, K_i be the closed curves surrounding edges to the right of Ω, as previously. Then, in addition to $\mathfrak{I}(\phi(w)) = 0$ on each edge of S, we have

$$(5)\qquad \begin{cases} \mathfrak{I}\left((w-\Omega)^{\frac{4\mu}{2\mu+1}}\phi(w)\right) \to 0 \text{ as } w \to \Omega \\ \mathfrak{R}\int_{C_j} \phi(w)\,dw = 0 \\ \mathfrak{R}\int_{K_i} \phi(w)\,dw = 0 \\ \mathfrak{I}\int_{N_\epsilon} \kappa(w)\phi(w)\,dw \to 0 \text{ as } N_\epsilon \to \Omega \end{cases}$$

where N_ϵ is a curve completely surrounding Ω in $\mathfrak{G}$, and $\kappa(w)$ is real on the boundary slit S. The last three of these relations (5) follow in exactly the same manner as (4). The first of these is obtained analogously to the second of the equations (1) by making the transformation

$$w - \Omega = \bar{w}^{2(\mu+\frac{1}{2})}.$$

This transformation maps the slit S to the line $\bar{v} = 0$, from which it follows that $\mathfrak{I}(\bar{\phi}(\bar{w})) = 0$ where

$$\bar{\phi}(\bar{w}) = (\mathfrak{x}_{\bar{u}} - i\mathfrak{x}_{\bar{v}})^2 = \phi(w)\left(\frac{dw}{d\bar{w}}\right)^2 = (2\mu + 1)^2 (w - \Omega)^{\frac{4\mu}{2\mu+1}} \phi(w).$$

In the neighborhood of $w = \Omega$, $\phi(w)$ is expandable into a power series in $(w - \Omega)^{\frac{1}{2\mu+1}}$ which is, by the first of the relations (5),

$$\phi(w) = \frac{A_{4\mu}}{(w-\Omega)^{\frac{4\mu}{2\mu+1}}} + \cdots + \frac{A_\mu}{(w-\Omega)^{\frac{2\mu}{2\mu+1}}} + \cdots + \frac{A_1}{(w-\Omega)^{\frac{1}{2\mu+1}}} + A_0 + \cdots$$

where $(w - \Omega)^{\frac{1}{2\mu+1}}$ is taken real on the boundary slit S, and the coefficients $A_{4\mu}, A_{4\mu-1}, \cdots, A_{2\mu}, \cdots, A_1, A_0, \cdots$ are all real. Taking

$$\kappa(w) = (w - \Omega)^{\frac{2\mu-1}{2\mu+1}},\ (w - \Omega)^{\frac{2\mu-2}{2\mu+1}}, \cdots, (w - \Omega)^{\frac{1}{2\mu+1}},\ 1$$

in the last of the relations (5) yields successively $A_{4\mu} = 0$, $A_{4\mu-1} = 0, \cdots$, $A_{2\mu+1} = 0$, so that $\mathfrak{R}\int_{w_0}^{w} \phi(w)\,dw$ has no singularity at $w = \Omega$.

The relations (4) hold for each inner branch point $w = W$ of $\mathfrak{G}$. From them, it follows that $\mathfrak{R}\int_{w_0}^{w} \phi(w)\,dw$ has no singularity at $w = W$. The first two of the equations (4) and the middle two of (5) show that $\mathfrak{R}\int_{w_0}^{w} \phi(w)\,dw$ is single valued in $\mathfrak{G}$. Exactly as in section 12, we have $\phi(w) \equiv 0$.

The proof of the main theorem is complete.

CONCLUDING REMARKS.

In the proof of the main theorem above, no use was made of the upper bound β of the inner diameters. Define d'_α as the greatest lower bound of $D(\mathfrak{x})$ for all allowable surfaces of inner diameter $\geqq \alpha$. Then, without further proof, we have the additional

THEOREM 2. *If $d'_\alpha < d_\alpha$ ($\alpha > 0$), there exists a minimal surface of the prescribed topological structure bounded by $\Gamma_1, \cdots, \Gamma_k$ of inner diameter $> \alpha$.*

Let A be the minimum of the diameters of the curves $\Gamma_1, \cdots, \Gamma_k$. The inner diameters of surfaces bounded by $\Gamma_1, \cdots, \Gamma_k$ range from 0 to A inclusive. The main theorem and Theorem 2 yield

THEOREM 3. *If d_α, considered as a function of α in $0 < \alpha \leqq A$, has a proper relative minimum for $\alpha = \bar{\alpha}$, there exists a minimal surface of the prescribed topological structure bounded by the curves $\Gamma_1, \cdots, \Gamma_k$ and having the inner diameter $\bar{\alpha}$.*

In addition to these, we have the following

APPROXIMATION THEOREM 4. *If $\Gamma_1^{(n)}, \cdots, \Gamma_k^{(n)}$, a sequence of closed Jordan curves approaching uniformly the Jordan curves $\Gamma_1, \cdots, \Gamma_k$, bounds a minimal surface $\mathfrak{x}^{(n)}$ (of given topological structure) of diameter $\alpha^{(n)}$, and if $D(\mathfrak{x}^{(n)}) < M$* and *$\alpha^{(n)} \to \alpha \neq 0$, then $\Gamma_1, \cdots, \Gamma_k$ bounds a minimal surface of inner diameter α of the prescribed topological structure.*

Proof. Let $\mathfrak{G}^{(n)}$ be the slit domain over which $\mathfrak{x}^{(n)}$ is defined. By virtue of $D(\mathfrak{x}^{(n)}) < M$ and $\alpha^{(n)} \to \alpha \neq 0$, it follows exactly as in the proof of the main theorem that there is a subsequence, renamed $\mathfrak{x}^{(n)}$, such that $\mathfrak{G}^{(n)}$ converges to a non-degenerate slit domain $\mathfrak{G}$ and $\mathfrak{x}^{(n)}$ converges uniformly to the potential surface $\mathfrak{x}(u, v)$. The derivatives of $\mathfrak{x}^{(n)}(u, v)$ converge to the respective derivatives of $\mathfrak{x}(u, v)$. The relations $\mathfrak{x}_u^{(n)2} = \mathfrak{x}_v^{(n)2}$, $\mathfrak{x}_u^{(n)}\mathfrak{x}_v^{(n)} = 0$ therefore imply $\mathfrak{x}_u^2 = \mathfrak{x}_v^2$, $\mathfrak{x}_u\mathfrak{x}_v = 0$, and $\mathfrak{x}(u, v)$ is a minimal surface.

The approximation theorem also holds if the restriction $D(\mathfrak{x}^{(n)}) < M$ is removed.

In the proof of all the Theorems 1, 2, 3, 4 above, no use has been made of the theory of conformal mapping. If it were used, the proof in sections 9-13 that $\phi(w) \equiv 0$ can be given in an extremely simple way, as has been shown by Courant.[28]

Furthermore, in all the above, essential use has been made of the condition that α be positive. It is possible to prove Theorems 1, 2 for the case when $\alpha = 0$, but it requires the theory of conformal mapping for domains of lower topological structure.[29] The inequality $d'_0 < d_0$ (Theorem 2, when $\alpha = 0$) then becomes the inequalities used by Courant and by Douglas.

Finally, the theory of conformal mapping yields the interpretation of the inequalities above in terms of area. By verifying these inequalities, the existence of minimal surfaces of higher topological structure can be established for large classes of bounding curves. For example, a knot will always bound a minimal surface of genus one and a non-orientable minimal surface; two interlocking curves will always bound at least two minimal surfaces.

NEW YORK UNIVERSITY.

---

[28] *Loc. cit.*, note 1, p. 717-720; or by the method of Radó.

[29] By using lower semi-continuity considerations similar to those of Courant, l. c.

UNITS IN p-ADIC ALGEBRAS.*

By O. F. G. Schilling.[1]

In this paper we shall discuss the unit groups of maximal orders in simple p-adic algebras. The fact that every such algebra is complete with respect to a suitably defined pseudo-valuation, leads to a tremendous simplification of our theory. It turns out that every unit of a given maximal order can be expressed as an infinite product of a finite number of basic units. The products are convergent with respect to the natural pseudo-valuation of the given simple algebra. The properties of the pseudo-valuation yield that the unit groups are totally disconnected locally compact groups.

Moreover, we can define the logarithmic and exponential functions in simple p-adic algebras. It can be shown that every unit which lies in a sufficiently small neighborhood of the group unit can be expressed as a value of the exponential function. Thus, the structure of our unit groups is similar to the structure of linear groups over the field of all real numbers. However, with regard to the associated Lie rings some differences show up. These discrepancies can be traced back to the fact that the pseudo-valuations under consideration satisfy a triangle inequality which is essentially stronger than the classical inequality. The p-adic integers play the rôle of all real numbers. Finally, it turns out that a sufficiently high power of every unit lies in the invariant subgroup which can be represented by values of the exponential function.

1. Units in division algebras. Let D be a division algebra of rank n over the field k of all p-adic numbers. Then the discrete valuation p of k has a uniquely determined extension P to D. Let o, O and $p = (\pi)$, $P = (\Pi)$ be the maximal orders and 2-sided prime ideals of k, D respectively. Then

$$pO = P^e, \quad [O/P : o/p] = f, \quad ef = n.$$

It is well known that O contains the $(p^f - 1)$-st roots of unity.

Suppose that $\{\omega_1, \cdots, \omega_f\}$ is a complete set of representatives of the residue field O/P over o/p. These representatives can be chosen to be the powers of a primitive root ω mod P.

Since D is complete with respect to the valuation P, every element $d \in D$ has an expansion

* Received March 8, 1939.

[1] Johnston Scholar at the Johns Hopkins University for the year 1938-1939.

$$d = \Pi^{\delta} \sum_{i=0}^{\infty} \omega_i \Pi^i$$

where $\omega_0 \neq 0$, $\delta \gtreqless 0$. If $\delta \geqq 0$ then $d \epsilon O$. In particular, d is a unit E of O if and only if $\delta = 0$.[2]

Let E be an arbitrary unit, then

$$E \equiv \omega^{\lambda} \pmod{P} \text{ where } 0 \leqq \lambda < p^f - 1.$$

We shall call H a 1-unit if

$$H \equiv 1 \pmod{P}.$$

Multiplying an arbitrary unit E by $\omega^{-\lambda}$ we get

$$E\omega^{-\lambda} \equiv \omega^{-\lambda}E \equiv \omega^{\lambda}\omega^{-\lambda} \equiv 1 \pmod{P}.$$

Hence $E = \omega^{\lambda}H_1 = H_2\omega^{\lambda}$ where H_1, H_2 are uniquely determined 1-units of O. Moreover, all 1-units H form an invariant subgroup of $\{E\}$ for

$$E^{-1}HE \equiv H \equiv 1 \pmod{P}.$$

Since the index $[\{E\} : \{H\}]$ is finite, the factor group $\{E\}/\{H\}$ being represented by the various powers ω^{λ}, it suffices to prove that $\{H\}$ has a finite base in order to be sure that $\{E\}$ has a finite base.

From now we shall be concerned with the group of 1-units $\{H\}$.[3] For every such 1-unit H we have

$$H = 1 + \omega_1\Pi + \omega_2\Pi^2 + \cdots.$$

DEFINITION 1. *Let*

$$H = 1 + \omega_h\Pi^h + \cdots, \ \omega_h \neq 0, \ i.e., \ H \equiv 1 \pmod{P^h}, \ \not\equiv 1 \pmod{P^{h+1}}.$$

Any such unit H is said to have degree h, $\deg H = h$.

DEFINITION 2. *The term $\omega_h\Pi^h$ of a unit H of degree h shall be called the principal term of H, $\mathfrak{P}(H)$.*

LEMMA 1. *If* $\deg(H_1) = \deg(H_2)$ *and* $\mathfrak{P}(H_1) = \mathfrak{P}(H_2)$, *then* $H_2 = HH_1 = H_1H'$, *where*

$$\deg H = \deg H' > h = \deg(H_1) = \deg(H_2), \textit{ and conversely.}$$

[2] H. Hasse, "Über p-adische Schiefkörper und ihre Bedeutung für die Arithmetik hyperkomplexer Zahlsysteme," *Mathematische Annalen*, vol. 104 (1931).

[3] The arguments used in the sequel go back to Hensel. See K. Hensel, "Die multiplikative Darstellung der algebraischen Zahlen für den Bereich eines beliebigen Primteilers," *Crelle*, vol. 146 (1916).

Proof. Let

$$H_1 = 1 + \omega\Pi^h + \omega_1\Pi^{h+1} + \cdots, \quad H_2 = 1 + \omega\Pi^h + \omega_2\Pi^{h+1} + \cdots.$$

Then

$$\begin{aligned} V_P(H_2H_1^{-1} - 1) &= V_P(H_2 - H_1)H_1^{-1} = V_P(H_2 - H_1) \\ &= V_P(1 + \omega\Pi^h + \omega_1\Pi^{h+1} + \cdots - 1 - \omega\Pi^h - \omega_2\Pi^{h+1} - \cdots) \\ &= V_P(\omega_1\Pi^{h+1} + \cdots - \omega_2\Pi^{h+1} - \cdots) = m > h. \end{aligned}$$

Thus,

$$\begin{aligned} H_2H_1^{-1} - 1 &= \omega_m\Pi^m + \cdots \quad \text{i.e.} \\ H_2H_1^{-1} &= 1 + \omega_m\Pi^m + \cdots = H, \\ H_2 &= HH_1 \text{ where } \deg H > h. \end{aligned}$$

The same argument yields the existence of a unit H' for which $H_2 = H_1H'$, $\deg H' = m > h$.

Conversely, let $H_2 = H_1H'$, e. g., where $\deg H' > h$. We want to prove $\mathfrak{P}(H_1) = \mathfrak{P}(H_2)$. Let $H_1 = 1 + \omega_1\Pi^h + \cdots$, $H' = 1 + \omega_3\Pi^{h+t} + \cdots$. Then

$$\begin{aligned} H_1H' &= 1 + \omega_1\Pi^h + \cdots + \omega_3\Pi^{h+t} + \cdots \\ &= 1 + \omega_1\Pi^h + \cdots = H_2 \text{ since } t \geqq 1. \end{aligned}$$

Consequently, $\mathfrak{P}(H_2) = \mathfrak{P}(H_1)$, for $\deg H_2 = \deg H_1$.

Let $\{H, h\}$ be the set of all units of degree h. This set is an invariant subgroup of finite index in the group $\{H\} = \{H, 1\}$ of all 1-units. Suppose that $H_i = 1 + \omega^{(i)}\Pi^h + \cdots$ are any 2 elements of $\{H, h\}$. Then

$$H_1H_2 = 1 + (\omega^{(1)} + \omega^{(2)})\Pi^h + \cdots, \quad H_2H_1 = 1 + (\omega^{(1)} + \omega^{(2)})\Pi^h + \cdots,$$

$$H_i^{-1} = 1 - \omega^{(i)}\Pi^h + \cdots.$$

Thus,

$V_P(H_1H_2^{-1}) = V_P(H_2H_1^{-1}) \geqq h$, $V_P(H_i^{-1} - 1) \geqq h$, i. e., H_1H_2, H_2H_1, H_i^{-1} all lie in $\{H, h\}$. Similarly, it follows that $H^{-1}H_iH \in \{H, h\}$ for any

$$H \in \{H, h - t\}, \; 0 < t < h - 1.$$

DEFINITION 3. *A set* $[H, h] = \{1 + \omega_i\Pi^h + \cdots\}$ *of* $\{H, h\}$ *which consists of* $p^f - 1$ *incongruent units* mod P^{h+1} *shall be called a base for the degree* h.

Remark. Let ω_i be a complete set of representatives of O/P over o/p; i. e., any ω can be expressed as $\Sigma\omega_ic_i$ where $0 \leqq c_i < p - 1$. Then the units $1 + \omega_i\Pi^h$ and their various powers constitute a base $[H, h]$. Moreover, every $H \in \{H, h\}$ is equal to a product $\prod_i (1 + \omega_i\Pi^h)^{c_i}$ to within units of $\{H, h + t\}$, $t \geqq 1$.

We state

LEMMA 2. *If* $H \in \{H, h\}$ *then* $H = H_h H'$ *where* $H_h \in [H, h]$ *and* $\deg H' \geqq h + 1$. *The unit* H' *is uniquely determined by* H *and the base* $[H, h]$.

Proof. Let $H = 1 + \omega\Pi^h + \cdots$, $\deg H = h$. There exists a uniquely determined unit $H_h \in [H, h]$ for which $\mathfrak{P}(H) = \mathfrak{P}(H_h)$. Then $H_h^{-1}H = H'$ where $\deg H' \geqq h + 1$ according to Lemma 1. Thus, $H = H_h H'$. Obviously, the unit H' is uniquely determined.

THEOREM 1. *Every unit* $H \in \{H\}$ *can be represented as an infinite* (*convergent*) *product*

$$H_{h_1}H_{h_1+h_2}\cdots H_{h_1+h_2+\cdots+h_i}\cdots \quad \textit{where}$$
$$\lim_{i=\infty}(h_1+\cdots+h_i)=\infty, \quad H_{h_1+\cdots+h_i}\in[H, h_1+\cdots+h_i].$$

Proof. Suppose that $\deg H = h_1$. Then, by Lemma 2, $H_{h_1}^{-1}H = H_1$ where $\deg H_1 = h_1 + h_2$, $h_2 \geqq 1$. Then $H_{h_1+h_2}^{-1}H_1 = H_2$, where $\deg H_2 = h_1 + h_2 + h_3$, $h_3 \geqq 1$; etc. Consider now the sequence of units

$$H_{h_1},\ H_{h_1}H_{h_1+h_2},\cdots,\ H_{h_1}H_{h_1+h_2}\cdots H_{h_1+h_2+\cdots+h_i},\cdots.$$

This sequence has a unique P-adic limit for

$$V_P(H_{h_1}\cdots H_{h_1+\cdots+h_i} - H_{h_1}\cdots H_{h_1+\cdots+h_{i-1}}) = h_1+\cdots+h_{i-1}\to\infty.$$

Since

$$H = H_{h_1}H_{h_1+h_2}\cdots H_{h_1+\cdots+h_i}H_i, \quad \text{where}$$
$$\deg H_i > h_1+\cdots+h_i$$

we have

$$H = H_{h_1}H_{h_1+h_2}\cdots H_{h_1+h_2+\cdots+h_i}\cdots.$$

The various factors $H_{h_1+\cdots+h_i}$ are uniquely determined by H and the bases $[H, h]$. Conversely, every infinite product is a unit of H.

Let $H = 1 + \omega\Pi^h + \cdots$, $\omega \not\equiv o \pmod P$. Then

$$H^p = 1 + p\omega\Pi^h + \cdots + \omega\Pi^h\cdots\omega\Pi^h + \cdots$$
$$= 1 + p\omega\Pi^h + \cdots + (\omega\Pi^h)^p + \cdots.$$

Thus,

$$V_P(H^p - 1) = e + h$$

provided that $e + h < hp$, i. e., $h > \dfrac{e}{p-1}$.

Let $p = E\Pi^e$, then

$$H^p = 1 + E\omega\Pi^{h+e} + \cdots.$$

Since E is a fixed unit, we see that $E\omega$ runs over the set of all residues mod P

if ω varies over the set of all residues distinct from 0. Consequently, $[H, h]^p$ is a base for $\{H, h+e\}$ if $h > \dfrac{e}{p-1}$. Denote by s the smallest positive integer contained in $\dfrac{e}{p-1}$. Then $[H, h]^p = [H, h+e]$ for $h = s+1, \cdots, s+e$. Thus,

$$[H, 1], \cdots, [H, s], [H, s+1], \cdots, [H, s+e],$$
$$[H, s+1]^p, [H, s+2]^p, \cdots, [H, s+i]^{p^j}, \cdots$$

where $0 \leqq i \leqq e$, $j = 1, 2, \cdots$, are complete bases for all degrees.

Since $\{H\}/\{H, s+e\}$ and $\{H, i\}/\{H, s+e\}$, $i < s+e$ are finite groups, we can pick representatives U, W of these factor groups lying in the various $\{H, i\}$ such that $U_1, \cdots, U_r$ form generators of the bases $[H, 2], \cdots, [H, s]$ and $W_1, \cdots, W_R$ form generators of the bases $[H, s+1], \cdots, [H, s+e]$. In general, the elements $U_1, \cdots, U_r, W_1, \cdots, W_R$ will not form a minimal set of generators of $\{H\}/\{H, s+e\}$. However, $W_1^{p^j}, \cdots, W_R^{p^j}$ constitute basic sets $[H, s+1]^{p^j}, \cdots, [H, s+e]^{p^j}$.

THEOREM 2. *Every unit H can be expressed as an infinite product*

$$H = \mathfrak{W}(U_1, \cdots, U_r)\mathfrak{W}_0(W_1, \cdots, W_R)$$
$$\times \mathfrak{W}_1(W_1^p, \cdots, W_R^p) \cdots \mathfrak{W}_j(W_1^{p^j}, \cdots, W_R^{p^j}) \cdots,$$

where $\mathfrak{W}(U_1, \cdots, U_r)$ is a finite product of the U's and the $\mathfrak{W}_j(W_1^{p^j}, \cdots, W_R^{p^j})$ are finite products $\prod_{l=1}^{R} W_l^{c_l p^j}$, $0 \leqq c_l \leqq p-1$.

Proof. Application of Lemma 2 and Theorem 1 yields that

$$H^* = [\mathfrak{W}(U_1, \cdots, U_r)\mathfrak{W}_0(W_1, \cdots, W_R)]^{-1}H$$

is a unit whose degree is $> s+e$. Since $\{H, h\}/\{H, h+1\}$ are finite abelian groups of type $(p, p, \cdots, p)$, the relation between $[H, s+1], \cdots, [H, s+e]$ and the basic sets of higher degree implies that H^* can be written in the form

$$H^* = \prod_{j=1}^{\infty} \mathfrak{W}_j(U_1^{p^j}, \cdots, U_R^{p^j}) = \prod_{j=1}^{\infty}\prod_{l} W_l^{c_{lj}p^j},$$

when $0 \leqq c_{lj} < p-1$. It is obvious that these infinite products are convergent.

Remark. Theorem 2 implies that $\{H\}$ has a finite base in the following sense. There exists a finite number of units $U_1, \cdots, U_r, W_1, \cdots, W_R$ in $\{H\}$ such that every element is an infinite convergent product of them. The representations which we obtain are of course not unique since the basic units chosen in the construction can satisfy non-trivial relations.

2. Units in general simple algebras. Let $A = D_m$ be an arbitrary simple algebra over the p-adic field k. The ring of all matrices of degree m with elements in O is a maximal order $O^* = \Sigma O c_{ik}$ of A. Moreover any other maximal order $\bar{O}$ of A is obtained from O^* by an inner automorphism of the algebra A, $a^{-1}O^*a = \bar{O}$ where a is a suitably chosen regular element of A. This isomorphism implies that the unit group E of $\bar{O}$ is isomorphic with unit group E^* of O^*. Thus, $\bar{E}$ and E^* have the same structure. In the sequel we shall consider E^*. In particular, we shall prove that E^* has a finite number of generators.

In order to construct a set of generators we shall make use of the fact that O is a principal ring and that A and D are complete algebras.

Let $K \geqq k$ be the center of A. Suppose that E is an arbitrary unit of O^*. Then $N_K E = \epsilon$, where N_K denotes the reduced norm of A with respect to the center K. Let $\epsilon = \zeta\eta$ where $\eta \equiv 1 \pmod{P \cap K}$; ζ a root of unity in K and $P \cap K$ the prime ideal of K. There exists a finite subgroup $\{\mathrm{Z}\}$ of E such that every root of unity in O^* is a transform of a unit in $\{\mathrm{Z}\}$. Moreover, according to the norm theory of p-adic algebras, $N_K\{\mathrm{Z}\} = \{\zeta\}$. Thus we can find for given $\zeta = \epsilon\eta^{-1}$ a unit $\mathrm{Z}(\epsilon) \in \{\mathrm{Z}\}$ such that $N_K \mathrm{Z}(\epsilon) = \zeta$. Hence

$$N_K \mathrm{Z}(\epsilon)^{-1}E = N_K E\mathrm{Z}(\epsilon)^{-1} = \zeta^{-1}\zeta\eta = \eta.$$

Thus, every element of $\{E\}$ can be obtained as $\mathrm{Z}H$ where $\mathrm{Z} \in \{\mathrm{Z}\}$ and $N_K H \equiv 1 \pmod{P \cap K}$. Thus, it suffices to prove that $\{H\}$ has a finite basis.

LEMMA 3. *The units $I_m + \lambda c_{ik}$ $(i \neq k)$ form a subgroup of $\{H\}$. The group $\{I + \lambda c_{ik}\} = G_{ik}$ is isomorphic with the additive group of O: I_m the unit matrix and $\lambda \in O$.*

Proof. Let $I_m + \lambda_j c_{ik}$ $(j = 1, 2)$ be any two elements of the set $I_m + \lambda c_{ik}$. Then, by matrix multiplication,

$$\begin{aligned}(I_m + \lambda_1 c_{ik})(I_m + \lambda_1 c_{ik}) &= I + \lambda_1 c_{ik} + \lambda_2 c_{ik} \\ &= I + (\lambda_1 + \lambda_2) c_{ik}.\end{aligned}$$

This equation immediately yields that the correspondence

$$I_m + \lambda c_{ik} \leftrightarrow \lambda$$

defines an isomorphism between $\{I + \lambda c_{ik}\} = G_{ik}$ and the additive group O. Thus G_{ik} is a group.

Next we show that G_{ik} has a finite base. Let $\alpha_1, \cdots, \alpha_n$ be a linearly independent base of O over o. Then every element $\lambda \in O$ has a unique representation $\lambda = \lambda_1\alpha_1 + \cdots + \lambda_n\alpha_n$ where $\lambda_i \in o$. Consequently,

$$I_m + \lambda c_{ik} = \prod_{l=1}^{n} (I_m + \lambda_j \alpha_j c_{ik}).$$

Observing that $\{E\}$ is a locally compact group[4] whose topology induces the topology of the locally compact group G_{ik} we get

$$I_m + \lambda c_{ik} = \prod_{l=1}^{n} (I_m + \alpha_j c_{ik})^{\lambda_j}.$$

Thus, the elements $I_m + \alpha_j c_{ik}$ $(j = 1, \cdots, n)$ form a base of G_{ik} if we admit infinite convergent products.

After these preliminaries we are now in a position to construct a finite base for $\{H\}$. We shall use arguments which are familiar in the theory of elementary divisors.

Let $H = (\beta_{ik})$, $\beta_{ik} \epsilon O$, $i, k = 1, \cdots, m$, be an arbitrary unit of $\{H\}$. Since the matrices T_m giving rise to the permutations of any pair of rows (or columns) of elements in A belong to $\{H\}$, we can proceed as follows. There exists a set of suitable permutation matrices T_m such that the element β_{11} of the resulting transform $H * \mathfrak{P}_m$ has minimal value. Or,

$$H * \mathfrak{P}_m = (\alpha_{ik}), \quad V_P(\alpha_{ik}) \geqq V_P(\alpha_{11}) \ (ik \neq 1), \quad \{\alpha_{ik}\} = \{\beta_{ik}\}.$$

Multiplying $H * \mathfrak{P}_m$ on the left and right by suitable elements g^*_1, g_1 of the various G_{ik}—addition of the 1st row (or column) to the others—we get

$$g_1 * (H * \mathfrak{P}_m) g_1 = \begin{pmatrix} \alpha_{11} & 0 \cdots 0 \\ 0 & \\ \cdot & \\ \cdot & B_{m-1} \\ 0 & \end{pmatrix}$$

where B_{m-1} is a square matrix of degree $m - 1$. Since all matrices used for this reduction belong to $\{H\}$, we have $g_1 * (H * P_m) g_1 \epsilon \{H\}$. Consequently, α_{11} is a unit ϵ_{11} of O and B_{m-1} is a unit E_{m-1} of matrix ring O_{m-1} of degree $m - 1$ over O.

Next we multiply $\begin{pmatrix} \epsilon_{11} & 0 \cdots 0 \\ 0 & \\ \cdot & \\ \cdot & E_{m-1} \\ 0 & \end{pmatrix}$ by the matrix $\begin{pmatrix} \epsilon_{11}^{-1} & 0 \cdots 0 \\ 0 & \\ \cdot & \\ \cdot & E_{m-1} \\ 0 & \end{pmatrix}$,

say on the right, and we get $\begin{pmatrix} 1 & 0 \cdots 0 \\ 0 & \\ \cdot & \\ \cdot & E_{m-1} \\ 0 & \end{pmatrix}$. The matrix E_{m-1} can now

[4] The proof that $\{E\}$ is locally compact shall be given later.

be multiplied by the matrix $\begin{pmatrix} 1 & 0 \cdots 0 \\ 0 & \\ \cdot & \\ \cdot & Z_{m-1} \\ 0 & \end{pmatrix}$ where $Z_{m-1} \epsilon \{Z_{m-1}\}$, $N_K\{Z_{m-1}\} = \{\zeta\}$ such that $E_{m-1}Z_{m-1}^{-1} = H_{m-1}$, $N_K H_{m-1} \equiv 1 \pmod{P \cap K}$. Thus we obtain a unit of the form $\begin{pmatrix} 1 & 0 \cdots 0 \\ 0 & \\ \cdot & \\ \cdot & H_{m-1} \\ 0 & \end{pmatrix}$. Now it is clear how we have to proceed.

If $n-1>2$, then we reduce H_{m-1} to $\begin{pmatrix} 1 & 0 \cdots 0 \\ 0 & \\ \cdot & \\ \cdot & H_{m-2} \\ 0 & \end{pmatrix}$ using the same arguments as before. Every such reduction can be performed in $\{E\}$. Namely, if S_{m-1} is matrix used for reduction of H_{m-1} then $\begin{pmatrix} 1 & 0 \cdots 0 \\ 0 & \\ \cdot & \\ \cdot & S_{m-1} \\ 0 & \end{pmatrix}$ is to be applied to the original unit; etc.

After a finite number of such transformations we obtain a matrix H_2 for which $N_K H_2 \equiv 1 \pmod{P \cap K}$. Permuting the elements of H_2 and subtracting the 1st row from the second, we can suppose that

$$H_2 = \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ 0 & \gamma_{22} \end{pmatrix}.$$

Then $H_2^{-1} = \begin{pmatrix} \gamma_{11}^{-1} & -\gamma_{11}^{-1}\gamma_{12}\gamma_{22}^{-1} \\ 0 & \gamma_{22}^{-1} \end{pmatrix}$. Hence γ_{11}, γ_{12} must be units of O. Put $H_2 = \begin{pmatrix} \epsilon_{11} & \gamma_{12} \\ 0 & \epsilon_{22} \end{pmatrix}$. Multiplication by $\begin{pmatrix} \epsilon_{11}^{-1} & 0 \\ 0 & 1 \end{pmatrix}$ yields

$$H_2 \begin{pmatrix} \epsilon_{11}^{-1} & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & \gamma_{12} \\ 0 & \epsilon_{22} \end{pmatrix}.$$

Next multiply by $\begin{pmatrix} 1 & -\gamma_{12} \\ 0 & 1 \end{pmatrix}$, we get $\begin{pmatrix} 1 & 0 \\ 0 & \epsilon_{22} \end{pmatrix}$. Finally multiplying by $\begin{pmatrix} 0 & \epsilon_{22}^{-1} \\ 1 & 0 \end{pmatrix}$ we obtain 1. Thus H_2 has a finite base since $\left\{ \begin{pmatrix} \epsilon_{11} & 0 \\ 0 & 1 \end{pmatrix} \right\} \cong E(O)$, $\left\{ \begin{pmatrix} 1 & 0 \\ 0 & \epsilon_{22} \end{pmatrix} \right\} \cong E(O)$ and $\left\{ \begin{pmatrix} 1 & \gamma \\ 0 & 1 \end{pmatrix} \right\} \cong O$. Every element of H_2 can be

expressed as an infinite (convergent) product of a finite number of units. (Observe Theorem 2 and Lemma 3.)

We state

THEOREM 3. *Every unit E of the maximal order O^* of $A = D_m$ can be expressed as an infinite convergent product of a finite number of units.*

Proof. As we have just seen $E = ZH$ where $N_K H \equiv 1 \pmod{P \cap K}$. Later we reduced H by elementary transformations. At every step of the reduction we obtained units which are convergent products of a finite number of basic units. The units being of the form

$$\begin{pmatrix} I_g & 0 \\ 0 & I_{m-g} + \alpha_j c_{i'k'} \end{pmatrix}, \quad 1 \leqq i',\ k' \leqq m-g,\ j = 1, \cdots, n \text{ and}$$

$$\begin{pmatrix} I_g & & & & & & 0 \\ & 1 & & & & & \\ & & \cdot & & & & \\ & & & \cdot & & & \\ & & & & \epsilon & & \\ & & & & & \cdot & \\ & & & & & & \cdot \\ 0 & & & & & & 1 \end{pmatrix}$$

where the ϵ are taken from a base of the unit group of D.

3. Topology in groups of units. In the preceding arguments we made frequent use of the fact that the various unit groups are topological groups. This fact enabled us to speak about convergent infinite products. To put these statements on a secure basis we prove

THEOREM 4. *The unit group of a maximal order O^* of $A = D_n$ is locally compact and complete.*

Proof. Let $O^* = O_m = \Sigma O c_{ik}$ be a maximal order of A. This maximal order contains a single 2-sided prime ideal $P^* = PO^*$ where P is the prime ideal of O. The prime ideal P^* gives rise to a pseudo-valuation V_{P^*} on A. Namely, let (a) be the intersection of all 2-sided ideals with respect to O^* which contain the element $a \in A$. Define

$V_{P^*}(a)$ as the least integer $\rho(a)$ such that

(a) $P^{*-\rho(a)}$ has a denominator which is relatively prime to P^*. It is easily seen that $\rho(a) = V_{P^*}(a)$ is a *pseudo-valuation* of A, i. e.,

(i) $V_{P^*}(a) < \infty$ if and only if $a \neq O$

(ii) $V_{P^*}(a) = V_{P^*}(-a)$

(iii) $V_{P^*}(a+b) \geqq \mathrm{Min}\,[V_{P^*}(a), V_{P^*}(b)]$

(iv) $V_{P^*}(ab) \geqq V_{P^*}(a) + V_{P^*}(b)$,

$V_{P^*}(ab) = V_{P^*}(a) + V_{P^*}(b)$ if $ab = ba$.[5]

It is clear that the pseudo-valuation V_{P^*} defines a topology on O^* such that the various integral powers P^{*i} are neighborhoods of the zero-element of A. Conversely, the chain

$$P^* \supset P^{*2} \supset P^{*3} \supset \cdots \supset P^{*i} \supset \cdots \underset{i}{\Delta} P^{*i} = 0$$

defines V_{P^*}. This equivalence yields, according to well-known principles of topology, that O^* is a totally disconnected locally compact group.

Define the distance between any two elements E_1, E_2 of the unit group $\{E\}$ of O^* as

$$\delta_{P^*}(E_1, E_2) = C^{V_{P^*}(E_1 - E_2)} \text{ where } 0 < C < 1.$$

It is obvious that δ_{P^*} satisfies the usual axioms of a distance function. Thus, $\{E\}$ is a metric group. Since the algebra A is complete with respect to the pseudo-valuation V_{P^*}, it follows that $\{E\}$ is complete with regard to the metric we just defined. As a consequence of property (iii) it follows that $\{E\}$ is totally disconnected. In order to see that $\{E\}$ is locally compact, we observe that the neighborhoods of 1 are given by the following congruences:

$$U_i(1) = \{\text{all } E_i \text{ such that } E_i \equiv 1 \pmod{P^{*i}}\}, \qquad (i = 1, 2, \cdots).$$

Moreover, we must observe that

$$O^*/P^{*i} \cong (O/P^i)_m$$

and

$$O^*/P^{*i+1} \to O^*/P^{*i} \cong O^*/P^{*i+1}/P^{*i}/P^{*i+1}.$$

These relations define a topology which is equivalent to the one given by the distance $\delta_{P^*}(E_1, E_2)$. Since the residue rings O^*/P^{*i} are finite, the associated subgroups of regular elements (they are the approximations of the units E) are finite, too. Consequently, E is locally compact. The fact that the algebra A is complete with respect to V_{P^*} yields that $\{E\}$ is complete.[6]

Remark 1. If we change the maximal order O^* to another maximal

---

[5] M. Deming, *Algebren*, Chap. IV, § 11 (Berlin, 1935); M. Moriya, "Zur Bewertung der einfachen Algebren," *Proceedings of the Imperial Academy of Japan*, vol. 13 (1937).

[6] D. van Dantzig, "Zur topologischen Algebra," *Mathematische Annalen*, vol. 107 (1933).

order $O^{**} = a^{-1}O^*a$ of A, the resulting pseudo-valuation $V_{P^{**}}$ is equivalent to V_{P^*}. Moreover, any pseudo-valuation of A is equivalent to V_{P^*}.

Remark 2. If A is a division algebra, then V_{P^*} is defined by the unique prime ideal of D. The pseudo-valuation V_{P^*} is a valuation in the customary sense, i. e., $V_{P^*}(ab) = V_{P^*}(a) + V_{P^*}(b)$ for all pairs of elements a, b in D.

Remark 3. The pseudo-valuation V_{P^*} of A induces a valuation in every division algebra which is contained in A. Consequently, V_{P^*} can be looked upon as the enveloping pseudo-valuation of all valuations on the subalgebras of A.

Remark 4. The various infinite products which we considered heretofore are convergent with respect to $\delta_{P^*}(E_1, E_2)$. In particular, the groups $\{I_m + \lambda c_{ik}\}$, $i \neq k$, are locally compact and complete. Their topology being the same as that of the imbedding group and the maximal order, respectively.

Since A contains the field of all p-adic numbers k and since

$$V_{P^*}(ab) = V_{P^*}(a) + V_{P^*}(b) \text{ if } ab = ba,$$

we can define the exponential and logarithmic functions as in the theory of p-adic fields. We define

$$(1) \quad \log a = \sum_{\nu=0}^{\infty} \frac{(-1)^{\nu-1}}{\nu} (a-1)^{\nu}$$

and

$$(2) \quad \exp b = \sum_{\nu=0}^{\infty} \frac{1}{\nu!} b^{\nu}.$$ [7]

LEMMA 4. *The logarithmic series* (1) *converges for all a for which $V_{P^*}(a-1) > 0$, $\log a$ is an element of A. It also converges for all a for which $(a-1)^{\rho} = 0$, $\rho < \infty$.*

Proof. The first statement of the Lemma is immediately reduced to the commutative case. Consider the ring $k[a]$. Since $k[a] \subset A$ it is a semi-simple commutative algebra. V_{P^*} induces a valuation or a pseudo-valuation on $k[A]$ depending on the fact whether the algebra $k[A]$ is a field or not. Since any such algebra is the direct sum of complete fields over k, we can apply the theory of the logarithmic function on a field. Thus, $\log H$ exists for units H which are $\equiv 1 \pmod{P^*}$.

In the second case the series $\log a$ has only a finite number of terms. Hence it necessarily is convergent.

Similar arguments yield:

[7] K. Hensel, *Zahlentheorie*, Berlin-Leipzig (1913).

LEMMA 5. *The exponential series* (2) *is convergent for every b for which either*

$$V_{P^*}(b) > \frac{1}{p-1} \quad \text{or} \quad B^\rho = 0, \qquad \rho < \infty$$

THEOREM 5. $a = \exp b$ *implies* $b = \log a$; *if* $a \equiv 1 \pmod{P^{*re}}$, $r > \dfrac{1}{p-1}$ *then*

$$b = \log a \text{ implies } b \equiv 0 \pmod{P^{*re}}$$

and

$$a = \exp b.$$

The proof of this theorem runs along the same lines as in the theory of p-adic fields.

These results can be used in order to obtain representations of all units $H_s \equiv 1 \pmod{P^{*s}}$, $s > \dfrac{1}{p-1}$, as values of the exponential function.

THEOREM 6. *Every unit H_s can be expressed as*

$$\exp b_s \text{ for some } b_s,\ s > \frac{1}{p-1}.$$

Proof. This follows immediately from Theorem 5. The arguments b_s form an ideal of O^*. Namely, the quantities b_s are determined by the property that their values are sufficiently large. It then follows from the theory of pseudo-valuations that they constitute a two-sided ideal of O^*. Using a base $\beta_1, \cdots, \beta_n$ of this ideal $B_s = \{b_s\}$ over o, we can say that every H_s has the form $\exp(\lambda_1\beta_1 + \cdots + \lambda_n\beta_n)$, $\lambda_i \in O$. Thus $\{H, s\} = \exp B_s$. This result corresponds to a well-known result in the theory of linear groups over the field of all real numbers.[8]

---

[8] For comparison see: J. von Neumann, "Uber die analytischer Eigenschaften von Gruppen linearer Transformationen und ihrer Darstellungen," *Mathematische Zeitschrift*, vol. 30 (1929); F. Hausdorff, "Die symbolische Exponentialformel in der Gruppentheorie," *Sachs. Gesellsch. der Wissensch. Leipzig. Berichte Math. Phys. Klasse*, vol. 58 (1906).

We want to point out that in our case the ring of all p-adic integers o takes the place of the field of all real numbers. It is possible to define the "infinitesimal group" of a unit group. However, the resulting enveloping ring admits only the elements of o as operators. The essential difference between the unit groups and groups of linear transformations can be sought in the fact that the exponential and logarithmic functions play a different role in both theories if one considers the regions of convergences. The set $\{\log H\}$ is in general not a Lie ring. However, $\{\log H_s\}$, for sufficiently large s, is a Lie ring which can be treated as in the case of linear groups. The inequality $V_{P^*}(a+b) \geqq \mathrm{Min}[V_{P^*}(a), V_{P^*}(b)]$ simplifies most of the considerations considerably.

THEOREM 7. *For every unit $H \equiv 1 \pmod{P^*}$ there exists a minimal exponent $\sigma(H)$ such that*

$$H^{\sigma(H)} \in \{H, h\}, \qquad h > 1.$$

Proof. We already observed that the groups $\{E\}$ and $\{H\}$ are complete with respect to the topology introduced by the homomorphisms

$$O^*/P^{*i+1} \to O^*/P^{*i} \cong O^*/P^{*i+1}/P^{*i}/P^{*i+1}, \qquad (i = 1, 2, \cdots).$$

Letting $\omega_{jk}^{(i)}$ be a fixed set of representatives of O/P and Π a prime element of O we find

$$a = \sum_{i=0}^{\infty} \Big(\sum_{j,K=1}^{m} c_{jk}\omega_{jk}^{(i)} \Big) \Pi^i$$

for every element $a \in A$. Thus, for every unit H_h which is $\equiv 1 \pmod{P^{*h}}$, $\not\equiv 1 \pmod{P^{*h+1}}$ we get

$$H_h = I_m + a_h$$

where

$$a_h = \sum_{i=h}^{\infty} \Big(\sum_{j,K=1}^{m} c_{jk}\omega_{jk}^{(i)} \Big) (I_m \Pi)^i.$$

The groups $\{H, h\}$ are invariant subgroups of all $\{H, h-t\}$, $t = 1, \cdots, h-1$, as follows immediately from the fact that V_{P^*} is a pseudo-valuation. Namely, let $H_{h-t} = 1 + a_{h-t}$, $H_h = 1 + a_h$. Then $H_{h-t}^{-1} = 1 - a_{h-t} + \cdots$. Thus,

$$\begin{aligned} H_{h-t}^{-1} H_h H_{h-t} &= (1 - a_{h-t} + \cdots)(1 + a_h)(1 + a_{h-t}) \\ &= 1 + a_h + a_{h-t}a_h a_{h-t} + a_{h-t}a_h - a_h a_{h-t} + \cdots. \end{aligned}$$

Consequently,

$$\begin{aligned} V_{P^*}(H_{h-t}^{-1} H_h H_{h-t} - 1) &= V_{P^*}(a_h + a_{h-t}a_h a_{h-t} + a_{h-t}a_h a_{h-t} + \cdots) \\ &\geqq \operatorname{Min}[V_{P^*}(a_h), V_{P^*}(a_{h-t}a_h a_{h-t}), V_{P^*}(a_{h-t}a_h), V_{P^*}(a_h a_{h-t})] \\ &\geqq \operatorname{Min}[V_{P^*}(a_h), 2V_{P^*}(a_{h-t}) + V_{P^*}(a_h), V_{P^*}(a_{h-t}) + V_{P^*}(a_h)] \\ &\geqq V_{P^*}(a_h). \end{aligned}$$

Thus, $H_{h-t}^{-1} H_h H_{h-t} \in \{H, h\}$ since $t \geqq 1$. It is obvious that the sets $\{H, h\}$ are groups. Now it is clear that $\{H\}$ is given as the limit of the factor groups $\{H\}/\{H, h\}$, $h = 2, 3, \cdots$. Since the groups $\{H\}/\{H, h\}$ are finite it follows that a sufficiently high power of every H_{h-t}, in particular of every H_1, lies in $\{H, h\}$.

COROLLARY 1. *If $A = D$ is a division algebra, then the $\sigma(H)$ are powers of p, since the indices $[\{H\} : \{H, h\}]$ are powers of p.*

COROLLARY 2. *Let $s = [1/(p-1)]$. Since $\{H\}/\{H, s+1\}$ is a finite group we can choose a finite set of representatives $H^{(1)}, \cdots, H^{(\lambda)}$ which are*

generators of that factor group. As a consequence every H can be expressed in the form

$$H = \mathfrak{N}(H^{(1)}, \cdots, H^{(\lambda)}) \exp b(H)$$

where $\mathfrak{N}(H^{(1)}, \cdots, H^{(\lambda)}) \bmod \{H, s+1\}$ is an element of $\{H\}/\{H, s+1\}$. The element $b(H)$ is determined by $\mathfrak{N}(H^{(1)}, \cdots, H^{(\lambda)})^{-1}H \in \{H, s+1\}$.

COROLLARY 3. *The groups $\{H, h\}$ also have finite bases. We only need to take the representations of a base of some $\{H, h\}/\{H, h+t\}$, t sufficiently large, and their p^j-th powers. The argument is the same as in the proof of Theorem 2 observing that $V_{P^*}(ab) \geqq V_{P^*}(a) + V_{P^*}(b)$.*

We next apply this corollary to the unit groups of arbitrary orders $\tilde{O}$ of A. Let $H(\tilde{O})$ be the group of units in $\tilde{O}$ for which $N_K H = 1$.

THEOREM 9. *The unit group $E(\tilde{O})$ of every order $\tilde{O}$ of maximal rank in A has a finite base.*

Proof. The given order $\tilde{O}$ is contained in at least one maximal order $a^{-1}O^*a$ of A. Without loss of generality we can suppose that $\tilde{O} \leqq O^*$ for $O^* \cong a^{-1}O^*a \supset a^{-1}E(O^*)a \supseteq E(\tilde{O})$. Consequently, there exists a least power $(I_m\Pi)^\tau$ of $I_m\Pi \in O^*$ such that

$$(I_m\Pi)^\tau O^* \subset \tilde{O}.$$

Thus, every unit of $\{H, \tau+\nu\}$, $\nu \geqq 0$, lies in $\tilde{O}$. Namely $H_{\tau+\nu} = 1 + a_{\tau+\nu} = 1 + (I_m\Pi)^\tau b$, $b \in O^*$, or $a_{\tau+\nu} \in \tilde{O}$. Since $H(\tilde{O}) \leqq H(O^*) = H$ and $\{H(O^*), \tau\} \leqq H(\tilde{O})$ we find that the indices $[H(\tilde{O}) : \{H(O^*), \tau\}]$ and $[H(O^*) : H(\tilde{O})]$ are finite. Application of the corollary for $h = \tau$ yields the theorem, for $[H(\tilde{O}) : \{H(O^*), \tau\}]$ is finite. From this one readily concludes that $E(\tilde{O})$ has a finite base. Similarly, it can be proved that $H(\tilde{O})$ contains a subgroup of finite index whose elements can be expressed as values of the exponential function.

THE JOHNS HOPKINS UNIVERSITY.

THE EXPONENTIAL REPRESENTATION OF CANONICAL MATRICES.*

By JOHN WILLIAMSON.

The present paper deals with certain questions on matrices, questions which, when applied to the complex field, concern the non-local structure of the complex group. From this point of view, the result, in case of the complex field, is to the effect that there exists in the group a set of elements which cannot be reached from the unit element by iteration of infinitesimal elements and is such as to have no interior elements. Actually the question will be approached directly in terms of matrices and the exceptional elements will be characterized explicitly. The situation is similar in case of the real subgroup of the complex group, a subgroup which is relevant in dynamics.[1] However the set of exceptional elements of this real subgroup does possess interior elements. This contradicts a lemma of K. Schröder,[2] which as a fact is also disproved by an example of E. Cartan.[3] It is also shown that each exceptional element is the product of two non-exceptional elements, one of which is of period two, thus verifying a general theorem of Cartan's.

Let G be the normal form of a real non-singular skew symmetric matrix,

$$G = \begin{pmatrix} 0 & E_n \\ -E_n & 0 \end{pmatrix},$$

where E_n is the unit matrix of order n. A matrix C which satisfies the equation

$$CGC' = G,$$

where C' is the transposed of the matrix C, has been called a canonical matrix. If S is an arbitrary symmetric matrix, a matrix C of the form

$$C = \exp(GS), \tag{1}$$

is canonical [1]; for

$$\begin{aligned} CGC' = \exp(GS)G\exp(S'G') &= \exp(GS)G\exp(-SG) \\ &= \exp(GS)\exp(-GS)G = G. \end{aligned}$$

* Received January 31, 1939.

[1] Aurel Wintner, "On linear conservative dynamical systems," *Annali di matematica pura ed applicata*, ser. 4, tomo 13 (1934-35).

[2] Kurt Schröder, "Einige Sätze aus der Theorie der kontinuierlichen Gruppen linearer Transformationen," *Dissertation*, Berlin (1934).

[3] Elie Cartan, "La théorie des groupes finis et continus et l'analysis situs," *Mémorial des sciences mathématiques*, fascicule 42 (1930), p. 21.

It is obvious, from (1), that the determinant of C is positive and therefore that it has the value plus one. However, though this is true for every canonical matrix,[4] not every canonical matrix has an exponential representation of the form (1). It is our purpose here to determine the nature of the canonical matrices, which do have such an exponential representation. While it is customary to restrict the term *canonical* matrix to real matrices, we shall also consider the case in which all matrices are matrices over the complex number field.

1. If D is a second canonical matrix, since the totality of canonical matrices forms a group, the matrix $D^{-1}CD$ is also canonical. When D and C are both canonical we shall say that the two canonical matrices C and $D^{-1}CD$ are *G-equivalent.* If D is a canonical and S a symmetric matrix, we shall say that the two symmetric matrices S and

$$T = D'SD, \tag{2}$$

are *G-congruent.* It should be noted that, since $G' = G^{-1}$, D' is also canonical.

Let C be a canonical matrix of the form (1) and let D be any canonical matrix. Then

$$D^{-1}CD = D^{-1}\exp(GS)D = \exp(D^{-1}GSD) = \exp(GD'SD) = \exp GT,$$

where T is given by (2). Hence we have

RESULT (*a*). *If C and F are two G-equivalent canonical matrices and $C = \exp(GS)$, then $F = \exp(GT)$, where S and T are G-congruent.*

In a similar manner it may be shown that we have

RESULT (*b*). *If two symmetric matrices S and T are G-congruent, the two canonical matrices* $\exp(GS)$ *and* $\exp(GT)$ *are G-equivalent.*

As a consequence of result (*a*) we see that, if a canonical matrix C has an exponential representation, so does every canonical matrix which is G-equivalent to C. Hence, in the discussion of our problem, we do not need to consider every canonical matrix C but simply one member, C_1, of each class of G-equivalent canonical matrices. Further, it follows from result (*b*) that we need only consider one member, S_1, of each class of G-congruent symmetric matrices. The matrix $\exp(GS_1)$ is a canonical matrix and may be taken as the matrix C_1 representative of its class of G-equivalent canonical matrices.

[4] John Williamson, "On the determinant of an automorph of a non-singular skew-symmetric matrix," *Bulletin of the American Mathematical Society*, vol. 45, No. 4, pp. 307-309 (April, 1939).

If a class of G-equivalent canonical matrices has no member which is of the form $\exp(GS_1)$, then no member of the class can have an exponential representation of the form (1). Since normal forms S_1 for symmetric matrices under G-congruent transformations as well as normal forms C_1 for canonical matrices under G-equivalent transformations have been determined,[5] it would be possible to solve our problem by comparing the matrices $\exp(GS_1)$ with the matrices C_1. Unfortunately the matrices S_1 and the matrices C_1, which have been obtained, are quite involved. The complications that would be thus introduced may be avoided to a large extent by the following considerations.[6]

Let C be a canonical matrix and let

$$G = PG_1P', \; C = PC_1P^{-1}. \tag{3}$$

Then C_1 satisfies

$$C_1G_1C'_1 = G_1. \tag{4}$$

Further, if F is a second canonical matrix and there exists a non-singular matrix Q, which satisfies

$$G = QG_1Q', \; F = QC_1Q^{-1}, \tag{5}$$

then C and F are G-equivalent; for

$$F = DCD^{-1}$$

where

$$D = QP^{-1} \text{ and}$$
$$DGD' = QP^{-1}G(P^{-1})'Q' = QG_1Q' = G,$$

by (3) and (5). Conversely, if C and F are G-equivalent and (3) is satisfied, we may determine a non-singular matrix Q such that (5) is true. Hence instead of considering a normal form for a canonical matrix C under G-equivalent transformations we may consider a *normal pair* G_1, C_1 defined by (3). For brevity we shall say that the pair G_1, C_1 is equivalent to the pair G, C. The normal pairs G_1, C_1 equivalent to any pair G, C, where C is canonical, have been determined,[7] and are comparatively simple in form.

If now, with the above notations,

$$C = \exp(GS), \; C_1 = P^{-1}\exp(GS)P = \exp(P^{-1}G(P^{-1})'P'SP) = \exp(G_1S_1),$$

[5] John Williamson, "On the algebraic problem concerning the normal forms of linear dynamical systems," *American Journal of Mathematics*, vol. 58 (Jan., 1936), pp. 141-163; "On the normal forms of linear canonical transformations in dynamics," *American Journal of Mathematics*, vol. 59 (July, 1937), pp. 599-617. These papers will be referred to as W_1 and W_2 respectively.

[6] Cf. W_1 and W_2. § 1.

[7] W_2. Results I-IV.

where

(6) $$S_1 = P'SP.$$

Since

$$G = PG_1P', \; G_1^{-1} = P'G^{-1}P,$$

so that the pair of matrices G^{-1}, S is congruent under the same transformation to the pair G_1^{-1}, S_1. Accordingly we may replace results (a) and (b) by the more general but more manageable,

RESULT (a_1). *Let the pair G_1, C_1 be equivalent to the pair G, C. If $C = \exp(GS)$, then $C_1 = \exp(G_1S_1)$, where the pair G_1^{-1}, S_1 is congruent to the pair G^{-1}, S;*

and

RESULT (b_1). *If the pair G_1^{-1}, S_1 is congruent to the pair G^{-1}, S the pair $G_1, \exp(G_1S_1)$ is equivalent to the pair $G, \exp(GS)$.*

As normal pairs G_1^{-1}, S_1 under congruent transformations are known, by considering in turn all matrices of the form $\exp(G_1, S_1)$, we shall determine all normal pairs G_1, C_1, where C_1 has an exponential representation of the form $\exp(G_1S_1)$. If C is canonical and the pair G, C is equivalent to the pair G_1, C_1, where $C_1 = \exp(G_1S_1)$, then C has an exponential representation of the form (1). If, on the other hand, the pair G, C is not equivalent to such a pair, C cannot have an exponential representation of the form (1).

Since the normal pairs G_1, C_1 and G_1^{-1}, S_1 depend on the field in which the elements of the matrices lie, it is no longer possible to consider the two cases of real matrices and complex matrices simultaneously. Since the complex field is algebraically closed, the results in this case are somewhat simpler than those in the real field. Accordingly we first consider the case in which all matrices are matrices over the field of all complex numbers.

2. The complex field. If C is a canonical matrix over the complex number field, the normal pair G_1, C_1 depends solely on the elementary divisors of $C - \lambda E$ or, as we shall say, on the elementary divisors of C. These elementary divisors are, however, not entirely arbitrary but are subject to the following restrictions: if $a \neq 1$ or -1 and $(\lambda - a)^r$ occurs exactly k times among the elementary divisors of C, then $(\lambda - a^{-1})^r$ occurs exactly k times among the elementary divisors of C; if $w = \pm 1$ and r is odd the elementary divisor $(\lambda - w)^r$ must occur an even number of times amongst the elementary divisors of C.

The matrices G_1 and C_1 of the normal pair are both similarly partitioned *diagonal block* matrices,

(7) $\qquad G_1 = [K_1, K_2, \cdots, K_s]$ and $C_1 = [R_1, R_2, \cdots, R_s]$,

where K_j is of the same order as R_j. The matrices K_j and R_j are of two types:

Type (i):

$$R_i = \begin{pmatrix} N_i & 0 \\ 0 & (N'_i)^{-1} \end{pmatrix}, \qquad K_i = \begin{pmatrix} 0 & E_i \\ -E_i & 0 \end{pmatrix},$$

where E_i is the unit matrix of the same order as N_i. The matrix N_i has the single elementary divisor $(\lambda - a)^r$ and, if r is even, a is not equal to one or minus one. The matrix R_i has therefore only the two elementary divisors $(\lambda - a)^r$ and $(\lambda - a^{-1})^r$. The matrix N_i may be replaced by any matrix similar to it but, if taken in the classical canonical form, is then uniquely determined.

Type (ii). The matrix R_i has the single elementary divisor $(\lambda - w)^{2r}$, where $w = \pm 1$.[8] We shall not further specify the matrices R_i and K_i as it is sufficient for our purpose that they exist and may be determined in an unique manner.

Since the matrices G_1 and C_1 in (7) are direct sums of matrices we may consider the single component matrices separately. Essentially we have the following: *For every positive integer r and for every complex number $a \neq 0$ there is one and only one normal pair G_1, C_1 of order $2r$, where the elementary divisors of C_1 are $(\lambda - a)^r$* and *$(\lambda - a^{-1})^r$; for every positive integer r there is one and only one normal pair C_1, G_1, where C_1 has the single elementary divisor $(\lambda - 1)^{2r}$ or $(\lambda + 1)^{2r}$.*

The matrices of the normal pair G_1^{-1}, S_1 congruent to G^{-1}, S are again similarly partitioned diagonal block matrices, the blocks depending in an unique manner on the elementary divisors of the pencil $S - \lambda G^{-1}$. These elementary divisors are subject to the following restrictions: if $p \neq 0$ and $(\lambda - p)^r$ occurs exactly k times among the elementary divisors, then so must $(\lambda + p)^r$; if r is odd, the elementary divisor λ^r must occur an even number of times.[9] More exactly

(8) $\qquad S_1 = [W_1, W_2, \cdots, W_k],\ G_1^{-1} = [H_1^{-1}, H_2^{-1}, \cdots, H_k^{-1}]$,

where the elementary divisors of $W_i - \lambda H_i^{-1}$ are of two types:

---

[8] W_2. Result III, page 611. Since the complex field is algebraically closed $\epsilon = 1$. The actual results quoted above may be obtained by suitable modifications from W_2, which deals only with the real field. They may also be obtained as particular cases from "Normal matrices over an arbitrary field of characteristic zero," *American Journal of Mathematics*, vol. 61 (April, 1939).

[9] H. W. Turnbull and A. C. Aitken, *Canonical Matrices*, Blackie and Sons, London, 1932, p. 125. W_1. Type *a* and type *b*.

Type (α). $(\lambda - p)^r, (\lambda + p)^r$;
Type (β). λ^{2r}.

In type (α), if r is even and $p = 0$, there is a slight duplication as this is then the same as two of type (β).

Since G_1 and S_1 are similarly partitioned diagonal block matrices, so is the matrix $\exp(G_1S_1)$. Therefore there is no loss in generality in treating each block separately or, what is equivalent to this, in considering the two special cases; that in which $S - \lambda G^{-1}$ has a single pair of elementary divisors of type (α) and that in which it has a single elementary divisor of type (β).

Type (α). If $S - \lambda G^{-1}$ has a single pair of elementary divisors of type (α),[10]

$$G_1^{-1} = G^{-1} = \begin{pmatrix} 0 & -E \\ E & 0 \end{pmatrix}, \qquad S_1 = \begin{pmatrix} 0 & N' \\ N & 0 \end{pmatrix},$$

where

$$N = pE + U$$

and E and U are respectively the unit matrix and the auxiliary unit matrix of order r. Therefore

$$\exp(G_1S_1) = \exp(-GS_1) = [\exp(N), \exp(-N')].$$

But

$$\exp(N) = \exp(p)\exp(U) = \exp(p)\,(E + U + U^2/2! + \cdots + U^{r-1}/(r-1)!).$$

Therefore $\exp(N)$ has the single elementary divisor $(\lambda - \exp(p))^r$. Hence, if $a = \exp(p)$, so that $p = \log(a)$, $\exp(G_1S_1)$ has the single pair of elementary divisors $(\lambda - a)^r$, $(\lambda - a^{-1})^r$. Since p is an arbitrary complex number, if $a \neq 0$, we can always determine p such that $p = \log a$. The matrix $\exp(G_1S_1)$, with the single pair of elementary divisors $(\lambda - a)^r$, $(\lambda - a^{-1})^r$ may be taken as the matrix C_1 of the normal pair G_1, C. This matrix C_1 is not, however, identical with the one previously obtained.[11] Incidentally it appears that this new normal form for C_1 might be more convenient. As a consequence of result (a) we have

LEMMA 1. *Every canonical matrix C, whose elementarry divisors are all of type* (i), *has an exponential representation of the form* (1).

Type (β). If $S - \lambda G^{-1}$ has the single elementary divisor λ^{2r}, $G_1^{-1} = FT$, $S_1 = FTU$,

[10] W_2, formulae (58) and (59).
[11] W_2. Result I.

where

(9) $$F = [1, -1, 1, -1, \cdots, 1, -1],$$

T is the counter unit matrix and U the auxiliary unit matrix of order $2r$.[12] Then $G_1S_1 = U$ and $\exp(G_1S_1) = \exp(U)$.

Since the matrix $\exp(U)$ has the single elementary divisor $(\lambda - 1)^{2r}$, we have

LEMMA 2. *A canonical matrix C with the single elementary divisor $(\lambda - 1)^{2r}$ has an exponential representation of the form* (1).

We have now considered all possible types of matrices $\exp(G_1S_1)$ and have not found one with the single elementary divisor $(\lambda + 1)^{2r}$. As our treatment has been exhaustive we are led to

LEMMA 3. *A canonical matrix C with the single elementary divisor $(\lambda + 1)^{2r}$ does not have an exponential representation of the form* (1).

Combining the results of Lemmas 1, 2 and 3, we obtain

THEOREM 1. *A canonical matrix C over the complex number field can be represented in the exponential form* (1), *if, and only if, no elementary divisor of the form $(\lambda + 1)^{2r}$ occurs an odd number of times amongst the elementary divisors of C.*

The simplest illustration of Lemma 3 is obtained from the two rowed matrix $\begin{pmatrix} -1 & -1 \\ 0 & -1 \end{pmatrix}$. This matrix is canonical, since the value of its determinant is unity, and it has the single elementary divisor $(\lambda + 1)^2$. It does not have an exponential representation of the form (1).

If C is a canonical matrix with the single elementary divisor $(\lambda + 1)^{2r}$, the normal pair C_1, G_1 equivalent to C, G may be taken as

$$C_1 = -\exp(U), \quad G_1 = FT,$$

where U, F and T are defined by (9).[13] If D is the diagonal block matrix

$$D = [d_1, d_2, \cdots, d_r, d_r^{-1}, \cdots, d_2^{-1}, d_1^{-1}],$$
$$DG_1D' = DFTD = FDTD = FTD^{-1}D = FT = G_1.$$

Since $DC_1 = -D\exp(U)$, the latent roots of DC_1 are $-d_i$, $-d_i^{-1}$, $i = 1, 2, \cdots, r$. Hence, if $d_i \neq 1$, no latent root of DC_1 is -1. But

---

[12] W_1. Page 156. Cf. type b when $p = U$; the matrices q and τ both have the value 1.

[13] This is not the normal pair previously found but it is a simpler one suggested by the results of this paper. That it is a normal pair follows from the facts that FT is skew-symmetric and that $FTU' = -UFT$.

$$DC_1G_1(DC_1)' = G_1$$

and, as no latent root of DC_1 has the value -1, by Theorem 1, DC_1 has an exponential representation of the form $\exp(G_1S_1)$. Since

$$\lim_{d_i \to 1} D = E, \quad \lim_{d_i \to 1} DC_1 = C_1.$$

Hence the matrix C_1, which has no exponential representation, may be obtained as the limit of matrices K satisfying $KG_1K' = G_1$, where each matrix K has an exponential representation. Therefore we have

THEOREM 2. *Every canonical matrix C, which cannot be represented in the exponential form* (1), *is the limit of canonical matrices which can be so represented.*

3. The real field. If C is a real canonical matrix, the elementary divisors of C are subject to the following restrictions. Let $(\lambda - a)^r$ occur exactly k times among the elementary divisors of C; then, if a is real and distinct from 1 or -1, $(\lambda - a^{-1})^r$ must occur exactly k times; if a is complex and $|a| \neq 1$, $(\lambda - \bar{a})^r$, $(\lambda - a^{-1})^r$ and $(\lambda - \bar{a}^{-1})^r$ must all occur exactly k times; if a is complex and $|a| = 1$, $(\lambda - \bar{a})^r$ must occur exactly k times; finally, if $a = \pm 1$ and r is odd k must be even.[14] The matrices G_1 and C_1 of the normal pair equivalent to the pair G, C under real transformations are again of the form (7). However, the matrices R_i and K_i, which are of course real, are not in every case uniquely determined by the elementary divisors of C. They are of the following types:

$$\textit{Type (i)}: \quad R_i = \begin{pmatrix} N & 0 \\ 0 & (N')^{-1} \end{pmatrix}, \qquad K_i = \begin{pmatrix} 0 & E \\ -E & 0 \end{pmatrix}.$$

Here N has the single elementary divisor $(\lambda - a)^r$, where a is real and, in case $a = \pm 1$, r is odd, or else N has the single pair of complex conjugate elementary divisors $(\lambda - a)^r$, $(\lambda - \bar{a})^r$, $|a| \neq 1$.

Type (ii). $\qquad R_i,\ K_i = \rho V_i, \qquad \rho = \pm 1,$

where R_i has the single pair of elementary divisors $(\lambda - a)^r$, $(\lambda - \bar{a})^r$ and $|a| = 1$.

Type (iii). $\qquad R_i,\ K_i = \rho V_i, \qquad \rho = \pm 1,$

where R_i has the single elementary divisor $(\lambda - \omega)^{2r}$, $\omega = \pm 1$. The ρ which occurs in types (ii) and (iii) has a value ± 1 and the pair R_i, V_i is not

[14] W_2, page 611.

equivalent to the pair $R_i, -V_i$. Apart from this the matrices R_i and K_i may, in all cases, be determined in an unique manner.[15]

The matrices of the normal pair G_1^{-1}, S_1 congruent to G^{-1}, S under real congruent transformations are again given by (8), where the matrices W_i and H_i are of course real. The elementary divisors of $W_i - \lambda H_i^{-1}$ are of several distinct types:

Type (α). $(\lambda - p)^r$, $(\lambda + p)^r$, p real;

Type (β). $(\lambda - p)^r$, $(\lambda + p)^r$, $(\lambda - \bar{p})^r$, $(\lambda + \bar{p})^r$, p complex and $p \neq -\bar{p}$.

Type (γ). $(\lambda - p)^r$, $(\lambda + p)^r$, p pure imaginary so that $p = -\bar{p}$,

Type (δ). λ^{2r}.

As in the complex case we now consider in turn a pencil $S - \lambda G^{-1}$ with elementary divisors of types (α), (β), (γ) and (δ) and determine the corresponding matrices $\exp(G_1 S_1)$.

Type (α). This is in every way similar to *type* (α) of the complex case and we have

$$S_1 = \begin{pmatrix} 0 & N' \\ N & 0 \end{pmatrix}, \qquad G_1^{-1} = -G, \qquad N = pE + U,$$

so that

$$\exp(G_1 S_1) = \begin{pmatrix} \exp(p)\exp(U) & 0 \\ 0 & \exp(-p)\exp(-U') \end{pmatrix}.$$

The elementary divisors of $\exp(G_1 S_1)$ are therefore $(\lambda - a)^r$, $(\lambda - a^{-1})^r$, where $a = \exp(p)$. Since p is a real number, a must be positive and we therefore have

LEMMA 4. *If a is a real positive number different from unity, a canonical matrix C with the single pair of elementary divisors $(\lambda - a)^r$, $(\lambda - a^{-1})^r$ has a real exponential representation of the form* (1).[16]

Type (β). Since p is complex and $p \neq -\bar{p}$, $p = c + id$, where c and d are real and c is not zero. We may without risk of confusion write p as a two rowed square matrix $p = \begin{pmatrix} c & d \\ -d & c \end{pmatrix} = c + id$, where $i = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. Then

$$S_1 = \begin{pmatrix} 0 & N' \\ N & 0 \end{pmatrix}, \qquad G_1^{-1} = -G, \qquad N = pE + U,$$

[15] W_2, page 611.

[16] Though this lemma is also true when $a = 1$, it has not yet been proved if r is even, for, when r is even, there are three non-equivalent normal pairs with the elementary divisors $(\lambda - 1)^r$, $(\lambda - 1)^r$. See Lemma 9.

where, in conformance with out notation, U is the direct product of the auxiliary unit matrix by the unit matrix of order two. Therefore

$$\exp(G_1S_1)=\begin{pmatrix}\exp(N) & 0\\ 0 & \exp(-N')\end{pmatrix}=\begin{pmatrix}\exp(p)\exp(U) & 0\\ 0 & \exp(-p')\exp(-U')\end{pmatrix}$$

But

$$\exp(p)=\exp(c)\begin{pmatrix}\cos d & \sin d\\ -\sin d & \cos d\end{pmatrix}$$

and

$$\exp(-p')=\exp(-c)\begin{pmatrix}\cos d & \sin d\\ -\sin d & \cos d\end{pmatrix}.$$

Hence, if $a=\exp(c+id)$, the elementary divisors of $\exp(p)\exp(U)$ are $(\lambda-a)^r$, $(\lambda-\bar{a})^r$ and those of $\exp(-p')\exp(-U')$ are $(\lambda-a^{-1})^r$, $(\lambda-\bar{a}^{-1})^r$. The elementary divisors of $\exp(G_1S_1)$ are therefore $(\lambda-a)^r$, $(\lambda-\bar{a})^r$, $(\lambda-a^{-1})^r$, $(\lambda-\bar{a}^{-1})^r$. Since c and d are arbitrary, except that $c\neq 0$, for any a, whose absolute value is not unity, we can always determine c and d such that $a=\exp(c+id)$. Hence we have

LEMMA 5. *If a is complex and $|a|\neq 1$, a real canonical matrix C with only the four elementary divisors $(\lambda-a)^r$, $(\lambda-a^{-1})^r$, $(\lambda-\bar{a})^r$, $(\lambda-\bar{a}^{-1})^r$ has a real exponential representation of the form* (1).

Further, if d is not an integral multiple of π, a is certainly complex; if d is an even multiple of π, a is real and positive; finally, if d is an odd multiple of π, $a=-\exp(c)$ and is negative. The elementary divisors of $\exp(p)\exp(U)$ are therefore, in this last case, $(\lambda+\exp(c))^r$, $(\lambda+\exp(c))^r$ and of $\exp(G_1S_1)$, $(\lambda+\exp(c))^r$, $(\lambda+\exp(c))^r$, $(\lambda+\exp(-c))^r$, $(\lambda+\exp(-c))^r$. Since c is arbitrary but not zero, we have

LEMMA 6. *If a is real and negative and different from minus one, a real canonical matrix whose only elementary divisors are $(\lambda-a)^r$, $(\lambda-a)^r$, $(\lambda-a^{-1})^r$, $(\lambda-a^{-1})^r$, has a real exponential representation of the form* (1).

It is important to notice that in Lemma 6, when a is negative, the elementary divisors $(\lambda-a)^r$ and $(\lambda-a^{-1})^r$ both occur twice.

Type (γ). With the notation used in the discussion of type (β) we let

$$p=\begin{pmatrix}0 & c\\ -c & 0\end{pmatrix}=ci \quad\text{and}\quad N=pE+U.$$

If ϵ is the unit matrix of order two and $F=[\epsilon,-\epsilon,\epsilon,\cdots,(-1)^{r-1}\epsilon]$, while T is the counter unit matrix of order r with 1 replaced by ϵ, then

(10) $G_1^{-1}=\rho TF$, when r is even, and $G_1^{-1}=\rho iTF$, when r is odd.[17]

---

[17] W_1. Case (3), page 162.

Since

$$FU = -UF \quad \text{and} \quad TU = U'T,$$

$$G_1^{-1}N = \rho TFN = (pE - U')\rho TF = -N'\rho TF = -N'G_1^{-1} = (G_1^{-1}N)'$$

Hence

$$S_1 = G_1^{-1}N \tag{11}$$

is symmetric.

The normal pair is therefore given by (10) and (11). There are two distinct normal pairs, one when $\rho = +1$ and the other when $\rho = -1$. Since

$$\exp(G_1S_1) = \exp(N) = \exp(p)\exp(U)$$

and

$$\exp(p) = \begin{pmatrix} \cos c & \sin c \\ -\sin c & \cos c \end{pmatrix},$$

the elementary divisors of $\exp(G_1S_1)$ are $(\lambda - a)^r$, $(\lambda - a^{-1})^r$, where $|a| = 1$. Since G_1 is only determined to within a factor $+1$ or -1, given by the value of ρ, we see that we thus obtain every canonical matrix of type (ii) and therefore have

LEMMA 7. *If a is complex and if $|a| = 1$, a real canonical matrix C with the single pair of elementary divisors $(\lambda - a)^r$, $(\lambda - \bar{a})^r$ has a real exponential representation of the form* (1).

If in the above $c = 2\pi$, $\exp(p) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, while, if $c = \pi$, $\exp(p) = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$. The elementary divisors of $\exp(G_1S_1)$ are then the pair $(\lambda - 1)^r$, $(\lambda - 1)^r$ or $(\lambda + 1)^r$, $(\lambda + 1)^r$. Since, when r is odd, an elementary divisor $(\lambda \pm 1)^r$ must occur an even number of times among the elementary divisors of C, and since the normal pair is then unique we deduce

LEMMA 8. *If r is odd and a real canonical matrix C has the single pair of elementary divisors $(\lambda - 1)^r$, $(\lambda - 1)^r$ or $(\lambda + 1)^r$, $(\lambda + 1)^r$, C has a real exponential representation of the form* (1).

We must examine the case, in which r is even, in more detail, since then (see type (iii)), if a canonical matrix C has the single pair of elementary divisors $(\lambda \pm 1)^r$, $(\lambda \pm 1)^r$, the normal pair is not unique. With each of the elementary divisors $(\lambda \pm 1)^r$ is associated a $\rho = \pm 1$. There are therefore three non-equivalent normal pairs corresponding to the values $+1, +1$; $-1, -1$; $+1, -1$ of the two ρ's that occur.[18] After a re-arrangement of

[18] W_2. Corollary to Theorem 1.

the rows and columns of all the matrices under consideration in the order $1, 3, 5, \cdots, r-1, 2, 4, \cdots, r$, $\exp(G_1S_1)$ reduces to

$$(12) \qquad \begin{pmatrix} \pm\exp(U) & 0 \\ 0 & \pm\exp(U) \end{pmatrix}$$

and G_1 to

$$(13) \qquad \begin{pmatrix} -\rho X & 0 \\ 0 & -\rho X \end{pmatrix}.$$

The matrix ρX is the matrix obtained from G_1^{-1} in (10) by replacing ϵ by 1. Its form is not important for us; the fact that the matrices in (12) and (13) are diagonal block matrices and that in (13) both blocks are exactly the same is, however, essential. Thus from type (γ) we cannot obtain a canonical matrix C with the two elementary divisors $(\lambda \pm 1)^r$, $(\lambda \pm 1)^r$ where one of the associated ρ's has the value $+1$ and the other the value -1.

Type (δ). Since r is even we may take [19]

$$G_1^{-1} = \rho TF, \quad \text{where} \quad \epsilon = 1,$$

and then

$$\exp(G_1S_1) = \exp(U).$$

The canonical matrix $\exp(G_1S_1)$ has the single elementary divisor $(\lambda - 1)^r$ and since G_1 has two possible values, corresponding to the two values 1 or -1 of ρ, we have

LEMMA 9. *If a real canonical matrix C has the single elementary divisor $(\lambda - 1)^{2k}$, it has a real exponential representation of the form* (1).

We have now exhausted all types (α), (β), (γ) and (δ) and on comparing the results obtained with the possible normal pairs of types (i), (ii) and (iii), we find that several simple types of canonical matrices do not have a real exponential representation. These types are:

(1) A canonical matrix with a single pair of real elementary divisors $(\lambda - a)^r$, $(\lambda - a^{-1})^r$ where a is negative and different from minus one;

(2) a canonical matrix with the single elementary divisor $(\lambda + 1)^{2k}$;

(3) a canonical matrix with the single pair of elementary divisors $(\lambda + 1)^{2k}$, $(\lambda + 1)^{2k}$, where in the normal pair one of the values of ρ is $+1$ and the other -1. This is a limiting case of (1) as a tends to -1.

The simplest example of type (3) is the following. If C has the two elementary divisors $(\lambda + 1)^2$, $(\lambda + 1)^2$ the three possible normal pairs equivalent to C, G are

---

[19] W_1, page 156.

$$\begin{pmatrix} j & 0 \\ 0 & j \end{pmatrix}, \begin{pmatrix} i & 0 \\ 0 & i \end{pmatrix}; \begin{pmatrix} j & 0 \\ 0 & j \end{pmatrix}, \begin{pmatrix} -i & 0 \\ 0 & -i \end{pmatrix}; \begin{pmatrix} j & 0 \\ 0 & j \end{pmatrix}, \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix},$$

where

$$j = \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix} \text{ and } i = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

In the first two cases C does have an exponential representation while in the third C does not.

We now combine our results in

THEOREM 3. *A real canonical matrix C has a real exponential representation of the form* (1) *if, and only if, every real elementary divisor of the form $(\lambda - a)^r$, where a is negative, occurs an even number of times and, when $a = -1$ and r is even, the number of positive ρ's associated with $(\lambda - a)^r$ is even* (i. e. *if $(\lambda + 1)^{2k}$ occurs $2m$ times among the elementary divisors of C, the index associated with this elementary divisor is even.*[20]

We now consider the canonical matrices, which do not have a real exponential representation and determine those which can be obtained as limiting cases of canonical matrices which do have such a representation. Let C have the single pair of elementary divisors $(\lambda - a)^r$, $(\lambda - a^{-1})^r$, where a is negative. Then

$$C_1 = [N_r, (N_r)^{-1}], \qquad G_1 = G,$$

where

$$N_r = aE + U.$$

If $r = 2k$, we note that

$$H_{2k} = \begin{pmatrix} a & 1 & 0 & 0 & . & 0 & 0 \\ -b^2 & a & 1 & 0 & . & 0 & 0 \\ 0 & 0 & a & 1 & . & 0 & 0 \\ 0 & 0 & -b^2 & a & . & 0 & 0 \\ . & . & . & . & . & 0 & 0 \\ 0 & 0 & 0 & 0 & . & a & 1 \\ 0 & 0 & 0 & 0 & . & -b^2 & a \end{pmatrix}$$

has only complex latent roots $a \pm ib$ when $b \neq 0$. The matrix $K_{2k} = [H_{2k}, (H'_{2k})^{-1}]$ is canonical and since, when $b \neq 0$, all latent roots of K_{2k} are complex, K_{2k} does have a real exponential representation. Since $\lim_{b\to 0} H_{2k} = N_{2k}$, $\lim_{b\to 0} K_{2k} = C_1$. Hence, if C is a canonical matrix with the single pair of elementary divisors $(\lambda - a)^r$, $(\lambda - a^{-1})^r$, where a is negative and r is even, C is the limit of canonical matrices which do have a real ex-

[20] W_2, page 613.

ponential representation. Further, when r is odd and equal to $2k+1$, the matrix

$$H_{2k+1} = \begin{pmatrix} H_{2k} & \gamma \\ 0 & a \end{pmatrix},$$

where $\gamma = (0, 0, \cdots, 0, 1)$ is a column vector of dimension $2k$, is such that $\lim_{b \to 0} H_{2k+1} = N_{2k+1}$.

When $b \neq 0$, the latent roots of H_{2k+1} consist of k pairs $a \pm ib$ and the single latent root a. Consequently, if C has only the four elementary divisors $(\lambda - a)^{2k+1}$, $(\lambda - a^{-1})^{2k+1}$, $(\lambda - a)^{2s+1}$, $(\lambda - a^{-1})^{2s+1}$, we may take

$$C_1 = [N_{2k+1}, N_{2s+1}, (N'_{2k+1})^{-1}, (N'_{2s+1})^{-1}] \quad \text{and} \quad G_1 = G.$$

Then, if $D = [H_{2k+1}, H_{2s+1}, (H'_{2k+1})^{-1}, (H'_{2s+1})^{-1}]$, D is canonical and $C_1 = \lim_{b \to 0} D$. When $b \neq 0$, the elementary divisors of D are all complex except for the four $(\lambda - a)$, $(\lambda - a)$, $(\lambda - a^{-1})$, $(\lambda - a^{-1})$. As a consequence of Lemmas 5 and 6, D has a real exponential representation.

Therefore, if C is a canonical matrix whose only latent roots are a and a^{-1} and if C is of order $4m$, so that a is a $2r$-fold root, C either has a real exponential representation or is the limit of canonical matrices which do. On the other hand, if C is of order $4m+2$, so that a is a $2r+1$-fold root and $C = \operatorname{Lim} M$, an odd number of the latent roots of M must be in the neighborhood of a and, if a is negative, must be negative. Hence M cannot have a real exponential representation.

If C has the single elementary divisor $(\lambda + 1)^{2k}$ we may take

$$C_1 = -\exp U, \qquad G_1 = \rho TF,$$

where F is defined by (9). If E is the unit matrix of order k and $H = [hE, h^{-1}E]$, the latent roots of HC_1 are $-h$ and $-h^{-1}$ both repeated k times. Since $HG_1H' = G_1$, the matrix HC_1 is the limit of matrices with real exponential representations when k is even. Since

$$\lim_{h \to 1} HC_1 = C_1,$$

the canonical matrix C, when k is even, is the limit of canonical matrices which have real exponential representations. When k is odd, we replace one of the h's in H by unity. Then the latent roots of HC_1 are $-h$, $-h^{-1}$ both repeated an even number of times and $-1, -1$.

The matrix

$$[-1, -1] = \exp\left\{ \begin{pmatrix} \pi & 0 \\ 0 & \pi \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \right\}.$$

The matrix

$$\begin{pmatrix} -1 & +1 \\ 0 & -1 \end{pmatrix} = \operatorname*{Lim}_{b\to 0} \frac{1}{\sqrt{1+b^2}} \begin{pmatrix} -1 & 1 \\ -b^2 & -1 \end{pmatrix} = \operatorname*{Lim}_{b\to 0} D,$$

where, since $|D| = 1$ and the latent roots of D are complex, D is canonical and has an exponential representation. Hence $HC_1 = \operatorname{Lim} W$, where $W = \exp(G_1S_1)$. Since $C_1 = \operatorname*{Lim}_{h\to 1} HC_1$, the canonical matrix C is also, in this case, the limit of matrices with a real exponential representation of the form (1). Combining these results we have

THEOREM 4. *A real canonical matrix C has a real exponential representation of the form* (1) *or is the limit of real canonical matrices which do, if, and only if, no negative number appears an odd number of times amongst the latent roots of C.*

Let C be a simple canonical matrix which does not have an exponential representation of the form (1). Then C either has a single elementary divisor $(\lambda+1)^{2r}$ or a pair of elementary divisors $(\lambda - a)^r$, $(\lambda - a^{-1})^r$, where a is negative. In either case C is of even order and, if E is the unit matrix of order $2r$, the canonical matrix $-E$ has, by Lemma 8, a real exponential representation of the form (1). Since the latent roots of $-C$ are the negatives of the latent roots of C, the canonical matrix $-C$ also has an exponential representation of the form (1). Further $-E$ is of period two and $C = -E(-C)$. As the general case of a canonical matrix, which does not have an exponential representation, may be reduced to the direct sum of simple canonical matrices we have proved,

THEOREM 5. *If C is a real canonical matrix, which does not have a real exponential representation, then C is the product of two real canonical matrices, one of period two, while both have a real exponential representation of the form* (1).

This is a particular case of a general theorem stated by Cartan.[21] It is obvious that (if the word real is omitted), Theorem 5 is valid in the complex field.

In a later paper the above methods will be used to discuss the similar problem, when G is replaced by an hermitian or symmetric matrix.

THE JOHNS HOPKINS UNIVERSITY.

[21] This fact was communicated to me by E. R. van Kampen.

METRIC METHODS IN DETERMINANT THEORY.* [1]

By Leonard M. Blumenthal.

Introduction. In previous papers the writer has applied results obtained in the study of the distance geometry of certain semimetric spaces to the theory of determinants.[2] The introduction of metric methods into this subject has given rise to new theorems, some of them being of quite a different *type* from those encountered in the classical theory. The application of abstract metrics to determinant theory has been useful not only in suggesting and proving theorems that without its motivation would hardly, perhaps, have been suspected, but use of such methods frequently succeeds in establishing whole chains of theorems rather than merely isolated results. This feature of the metric *Program,* applied to determinant theory, is instanced in several of the theorems developed in this note.

The present paper is concerned with the behavior of certain types of symmetric determinants when some of their principal minors are subjected to various conditions. Of the chains of theorems proved here, some were conjectured by the writer in the earlier papers referred to above, but their proofs awaited the completion of investigations in the distance geometry of pseudo-spherical sets of points. The results of these investigations (recently brought to a close) are now available.[3] In addition to their utilization, attention is directed to the geometric device employed in Theorems 3.1 and 4.1, and to the unsolved problems that it suggests.[4]

1. Determinant form of characterization theorem for pseudo-spherical sets. A semimetric space is formed by attaching to each pair of elements (*points*) p, q of an abstract set a non-negative real number (*distance*) pq, independent of order, such that $pq = 0$ if and only if $p = q$. A fundamental

* Received February 6, 1939.

[1] Presented to the Society, April 14, 1939.

[2] *Bulletin of the American Mathematical Society,* vol. 37 (1931), pp. 752-758; vol. 38 (1932), pp. 283-288; *American Journal of Mathematics,* vol. 56 (1934), pp. 225-232; *Duke Mathematical Journal,* vol. 2 (1936), pp. 396-404. See also Chapter IV of the author's "Distance geometries," *University of Missouri Studies,* vol. 13 (1938).

[3] L. M. Blumenthal and G. R. Thurman, "The characterization of pseudo-$S_{n,r}$ sets," *Proceedings of the National Academy of Sciences,* vol. 24 (1938), pp. 557-558. A brief summary of results is given.

[4] It is clear that instead of phrasing the applications of distance geometry dealt with in this note in the language of determinant theory, they may equally well be expressed as theorems in matrices, quadratic forms, etc.

problem consists in characterizing metrically those semimetric spaces Σ possessing the following properties: (1) if $S_{n,r}$ denotes the n-dimensional spherical surface with space constant r (the surface of a sphere of radius r in euclidean space of $n+1$ dimensions, with "shorter arc" distance), then corresponding to each set of $n+2$ points of Σ there is a function that maps the set upon the $S_{r,r}$ *with preservation of distances* (i. e., *congruently*); (2) the space Σ cannot be mapped congruently upon $S_{n,r}$; (3) if $p, q \in \Sigma$, then $pq \neq d = \pi r$, (i. e., no two points of Σ are *diametral*); (4) Σ contains more than $n+3$ points. The solution of this problem is to be applied in this paper to establish the new determinant theorems with which it is concerned. But first, it is necessary to put the solution in a form convenient for our purposes.

A semimetric space Σ satisfying the first two of the above conditions (and hence containing at least $n+3$ points) is called a pseudo-$S_{n,r}$ set. It is remarked that $n+2$ is the greatest integer for which (1) and (2) are compatible, for congruence of Σ with a subset of $S_{n,r}$ follows from the congruent imbedding in $S_{n,r}$ of each set of $n+3$ points of Σ.[5] We shall need the following theorems, the proofs of which will be published elsewhere.[6]

THEOREM 1. 1. *If the $n+3$ points $p_1, p_2, \cdots, p_{n+2}, p_{n+3}$ form a pseudo-$S_{n,r}$ set, then each set of $n+1$ of these points is an independent set (i. e., if $p_{i_1}, p_{i_2}, \cdots, p_{i_{n+1}}$ are any $n+1$ points contained in $p_1, p_2, \cdots, p_{n+3}$, then the determinant $|\cos(p_{i_j}p_{i_k}/r)|$, $(j, k = 1, 2, \cdots, n+1)$, does not vanish).*

THEOREM 1. 2. *If the two sets $p_1, p_2, \cdots, p_{n+2}, p_{n+3}$ and $q_1, q_2, \cdots, q_{n+2}, q_{n+3}$ form pseudo-$S_{n,r}$ sets such that*

$$p_1, p_2, \cdots, p_{n+2} \approx q_1, q_2, \cdots, q_{n+2},^{7}$$

then either $p_i p_{n+3} = q_i q_{n+3}$, $(i = 1, 2, \cdots, n+2)$, and the two sets are congruent, or $p_i p_{n+3} + q_i q_{n+3} = d$, $(i = 1, 2, \cdots, n+2)$.

THEOREM 1. 3. *If a pseudo-$S_{n,r}$ set of $n+4$ points contains no pair of diametral points, then each set of $n+3$ of its points forms a pseudo-$S_{n,r}$ set.*

COROLLARY. *If a pseudo-$S_{n,r}$ set of $n+4$ points contains no pair of diametral points, then any two of the $\frac{1}{2}(n+3)(n+4)$ distances determined by the $n+4$ points are either equal or their sum equals d.*

This Corollary is an immediate consequence of applying Theorem 1. 2 to pairs of the $n+4$ sets of $(n+3)$-tuples contained in the pseudo-$S_{n,r}$ set of

[5] "Distance geometries," p. 61.
[6] Proofs of these theorems for $n = 2$ are given in "Distance geometries," pp. 76-80.
[7] This notation signifies that $p_i p_j = q_i q_j$, $(i, j = 1, 2, \cdots, n+2)$.

$n+4$ points, each one of these $(n+3)$-tuples forming, according to Theorem 1.3, a pseudo-$S_{n,r}$ set. Clearly, $0 < p_i p_j < d$, $(i, j = 1, 2, \cdots, n+4; i \neq j)$.

Consider, now, the symmetric determinant

$$\Delta_{n+4}(p_1, \cdots, p_{n+4}) = |\cos(p_i p_j / r)|,$$

$(i, j = 1, 2, \cdots, n+4)$, of order $n+4$, formed for the $n+4$ pairwise distinct points $p_1, p_2, \cdots, p_{n+4}$ (no pair diametral) constituting a pseudo-$S_{n,r}$ set. By the above Corollary, any two elements of this determinant, no one of which is contained in the principal diagonal, *differ at most in sign.* Since each $n+2$ of the points $p_1, p_2, \cdots, p_{n+4}$ is congruent with $n+2$ points of the $S_{n,r}$, it follows that each principal minor of order $n+2$ of Δ_{n+4} vanishes, and hence no element of the determinant is zero.[8] Also, since $p_i \neq p_j$, $(i \neq j)$, $\cos(p_i p_j / r) \neq 1$, $(i \neq j)$.

REMARK. *Apart from the elements in the first and the j-th rows,* $(j = 2, 3, \cdots, n+4)$, *either each element in the first column of* Δ_{n+4} *is equal to the corresponding element in the j-th column, or each element in the first column is the negative of the corresponding element in the j-th column.*

Consider, for example, the first and second columns. According to Theorem 1.3, the two $(n+3)$-tuples $p_1, p_3, \cdots, p_{n+3}, p_{n+4}$; $p_2, p_3, \cdots, p_{n+3}, p_{n+4}$ are pseudo-$S_{n,r}$ sets. Since they obviously contain congruent $(n+2)$-tuples, it follows from Theorem 1.2 that either $p_i p_1 = p_i p_2$, $(i = 3, 4, \cdots, n+4)$, or $p_i p_1 + p_i p_2 = d$, $(i = 3, 4, \cdots, n+4)$. The same considerations applied to the first and the j-th columns verifies the remark in general.

A close examination of the determinant Δ_{n+4}, in the light of the foregoing observations, enables one to evaluate each of its elements (to within sign), and to describe the possible distributions of the signs. For suppose that the first column of the determinant contains, in addition to the element 1, exactly p positive elements, $0 \leqq p \leqq n+3$. If $p = 0$, then the elements $\cos(p_1 p_j / r)$ are negative, $(j = 2, 3, \cdots, n+4)$, and from the above Remark, either the elements $\cos(p_2 p_j / r)$ are *all* negative, or they are *all* positive, $(j = 3, 4, \cdots, n+4)$. In the first case, it follows from the symmetry of the determinant (and the Remark) that *every* element outside the principal diagonal is negative. Then the vanishing of the $(n+2)$-nd order principal minor $\Delta_{n+2}(p_1, \cdots, p_{n+2})$ yields at once that $\cos(p_i p_j / r) = -1/(n+1)$, $(i, j = 1, 2, \cdots, n+4; i \neq j)$.[9] In the second case, the same procedure

---

[8] "Distance geometries," p. 73. Since for no pair of indices i, j does $\cos(p_i p_j / r)$ vanish, $p_i p_j \neq d/2$, for every i, j.

[9] Denoting the determinant $\Delta_{n+2} = |r_{ij}|$, $r_{ij} = r_{ji} = \rho$, $(i \neq j)$, $r_{ii} = 1$ by $\Delta_{n+2}(\rho)$, it is found that $\Delta_{n+2}(\rho) = (1-\rho)^{n+1} \cdot [1 + (n+1)\rho]$.

shows that all the elements $\cos(p_ip_j/r)$ are positive, $(i, j = 2, 3, \cdots, n+4;\ i \neq j)$. But this is impossible, for the principal minor $\Delta_{n+2}(p_3, p_4, \cdots, p_{n+4})$ of order $n+2$ then fails to vanish. Hence the first case alone is possible, and it may be concluded that if $p = 0$, then each element of Δ_{n+4} outside the principal diagonal has the value $-1/(n+1)$.

There is no difficulty in applying the same procedure to the general case. The following two cases, represented schematically, are obtained:

Case A   *Case B*

$$\overbrace{\qquad\qquad}^{p+1} \qquad\qquad\qquad\qquad \overbrace{\qquad\qquad}^{p+1}$$

$$\begin{vmatrix} 1 & + & + & . & . & + & - & - & . & . & - \\ + & 1 & - & . & . & - & + & + & . & . & + \\ + & - & 1 & . & . & - & + & + & . & . & + \\ . & . & . & . & . & . & . & . & . & . & . \\ + & - & - & . & . & 1 & + & + & . & . & + \\ - & + & + & . & . & + & 1 & - & . & . & - \\ - & + & + & . & . & + & - & 1 & . & . & - \\ . & . & . & . & . & . & . & . & . & . & . \\ - & + & + & . & . & + & - & - & . & . & 1 \end{vmatrix}, \qquad \begin{vmatrix} 1 & + & + & . & . & + & - & - & . & . & - \\ + & 1 & + & . & . & + & - & - & . & . & - \\ + & + & 1 & . & . & + & - & - & . & . & - \\ . & . & . & . & . & . & . & . & . & . & . \\ + & + & + & . & . & 1 & - & - & . & . & - \\ - & - & - & . & . & - & 1 & + & . & . & + \\ - & - & - & . & . & - & + & 1 & . & . & + \\ . & . & . & . & . & . & . & . & . & . & . \\ - & - & - & . & . & - & + & + & . & . & 1 \end{vmatrix}$$

In Case A, multiplication of the 2nd, 3rd, $\cdots$, $(p+1)$-st rows and coluumns by -1 makes every element outside the principal diagonal negative, and we find, as before, that each such element has the value $-1/(n+1)$. Case B is seen to be impossible, for multiplication of the last $n-p+3$ rows and columns by -1 makes all the elements of the determinant positive, and the same contradiction as before is encountered.

The results thus obtained are stated in the form of the theorem:

THEOREM 1.5. *If P is a pseudo-$S_{n,r}$ set of $n+4$ points, $p_1, p_2, \cdots, p_{n+4}$, no two of which are diametral, then for every pair of distinct points p_i, p_j of P, $\cos(p_ip_j/r) = \pm 1/(n+1)$. The plus and minus signs are so distributed that the determinant of the $n+4$ points*

$$\Delta_{n+4}(p_1, p_2, \cdots, p_{n+4}) = |\cos(p_ip_j/r)|, \qquad (i, j = 1, 2, \cdots, n+4),$$

may upon multiplication of appropriate rows and the same numbered columns by -1, be transformed into a determinant $\Delta_{n+4}(-1/n+1)$ with each element outside the principal diagonal equal to $-1/(n+1)$.

There is no difficulty in extending Theorem 1.5 to pseudo-$S_{n,r}$ sets of arbitrary power *exceeding* $n+3$. The following theorem is obtained:

THEOREM 1.6. *If P is a pseudo-$S_{n,r}$ set containing more than $n+3$ points and without diametral points, then for every pair of distinct points p, q of P, $\cos(pq/r) = \pm 1/(n+1)$. For each positive integer k, the determinant*

$\Delta_{k+1}(p_1, p_2, \cdots, p_{k+1})$, *formed for* $k+1$ *points of* P *has, upon multiplication of appropriate rows and the same numbered columns by* -1, ALL *elements outside the principal diagonal equal to* $-1/(n+1)$.

This theorem gives the metric characterization of pseudo-$S_{n,r}$ sets containing more than $n+3$ points, and without diametral points, in a form suitable for our purposes. We proceed to develop some consequences of it.[10]

2. Determinants of type $\Delta = |r_{ij}|$, $r_{ij} = r_{ji}$, $-1 < r_{ij} < 1$, $(i \neq j)$, $r_{ii} = 1$, $(i, j = 1, 2, \cdots, m)$. An immediate consequence of Theorem 1.6 (together with earlier theorems on the metric characterization of the $S_{n,r}$)[11] is the theorem:

THEOREM 2.1. *Let* Δ *be of order* $m > n+3$, *where* n *is any given positive integer. If* (i) *every principal minor of order less than* $n+2$ *is nonnegative,* (ii) *every principal minor of order* $n+2$ *vanishes,* (iii) *at least one principal minor of order* $n+3$ *does not vanish, then* (1) *upon multiplying appropriate rows and the same numbered columns of* Δ *by* -1, *each element outside the principal diagonal has the value* $-1/(n+1)$, *and* (2) *for each positive integer* k, $(1 \leqq k \leqq m)$, *and each* k*-th order principal minor* Δ_k *of* Δ,

$$\Delta_k = \frac{1}{n+1}\left[\frac{n+2}{n+1}\right]^{k-1} \cdot (n-k+2).$$

Proof. Setting $r_{ij} = \cos \alpha_{ij}$, $0 < \alpha_{ij} < \pi$, $(i, j = 1, 2, \cdots, m;\ i \neq j)$, introduce a semimetric set of m points $p_1, p_2, \cdots, p_m$ with distances defined by $p_ip_j = r \cdot \alpha_{ij}$, $r > 0$, $(i, j = 1, 2, \cdots, m;\ i \neq j)$. Then

$$\Delta = \Delta_m(p_1, p_2, \cdots, p_m) = |\cos(p_ip_j/r)|, \qquad (i, j = 1, 2, \cdots, m).$$

Since $p_ip_j < \pi r$ for each pair of indices i, j, hypotheses (i), (ii) imply that each set of $n+2$ points contained in the m points is congruent with $n+2$ points of the $S_{n,r}$, while from (iii) it follows that the m points contain at least one set of $n+3$ points that is not congruent with a subset of $S_{n,r}$. Hence $p_1, p_2, \cdots, p_m$ are not embeddable isometrically in the $S_{n,r}$, and the set forms, consequently, a pseudo-$S_{n,r}$ set of the kind characterized in the preceding section. Then, by Theorem 1.6, the determinant Δ takes the form $\Delta_m(-1/n+1)$ upon multiplying appropriate rows and the same numbered columns by -1. The value of Δ_k, $1 \leqq k \leqq m$, given in the theorem, is now obtained by an easy computation.

---

[10] For the developments that follow it suffices that the pseudo-$S_{n,r}$ set P of Theorem 1.6 be composed of a finite number m of points, with $m > n+3$.

[11] *Loc. cit.*, footnote 7.

Thus, the conditions (i), (ii), (iii) placed upon those principal minors of Δ having orders not exceeding $n+3$ are sufficient to determine (essentially) the value of *every element* of Δ with arbitrary order exceeding $n+3$. The condition that $m > n+3$ is essential for the validity of the theorem.

3. Determinants of type $\Delta_N = | r_{ij} |$, $r_{ij} = r_{ji}$, $-1 < r_{ij} \leqq 0$, $(i \neq j)$, $r_{ii} = 1$, $(i, j = 1, 2, \cdots, m)$. A theorem somewhat stronger than Theorem 2.1 can be proved for determinants of type Δ_N under *weaker* hypotheses than those made in the preceding theorem.

THEOREM 3.1. *Let Δ_N be of order $m > n+3$. If* (i) *every principal minor of order less than $n+1$ is non-negative, while every principal minor of order $n+1$ is positive,* (ii) *every principal minor of order $n+2$ vanishes, then every element of Δ_N outside the principal diagonal has the value $-1/(n+1)$, and*

$$\Delta_k = \frac{1}{n+1}\left[\frac{n+2}{n+1}\right]^{k-1} \cdot (n-k+2), \qquad 1 \leqq k \leqq m.$$

Proof. Comparing the hypotheses of this theorem with those of Theorem 2.1, it is observed that (i) is slightly stronger than before, while hypothesis (iii) of Theorem 2.1 *does not appear at all.* To prove the theorem, it evidently suffices to show that for determinants of type Δ_N, hypotheses (i), (ii) imply that at least one principal minor of the determinant, of order $n+3$, does not vanish.

Introduce, as before, a semimetric set of m points $p_1, p_2, \cdots, p_m$, and put $r_{ij} = \cos(p_i p_j / r)$, $r > 0$, $(i, j = 1, 2, \cdots, m)$. Since $-1 < r_{ij} \leqq 0$ for each pair of distinct indices i, j, it may be assumed that $\frac{1}{2}\pi r \leqq p_i p_j < \pi r$, $(i, j = 1, 2, \cdots, m; i \neq j)$. We make the assumption that every $(n+3)$-rd order principal minor of $\Delta_N = \Delta_m(p_1, p_2, \cdots, p_m)$ vanishes, and show that this assumption leads to a contradiction.

It follows from (i), (ii) and the above assumption, that every set of $n+3$ points contained in the m points $p_1, p_2, \cdots, p_m$ may be embedded congruently in the $S_{n,r}$, and hence the whole set of m points is congruent with a subset of the $S_{n,r}$.[12] Thus, the $S_{n,r}$ must contain $m > n+3$ points $s_1, s_2, \cdots, s_m$ such that $s_1, s_2, \cdots, s_m \approx p_1, p_2, \cdots, p_m$, and hence the m points $s_1, s_2, \cdots, s_m$ have pairwise distances greater than or equal to $\frac{1}{2}\pi r$ and less than πr. From the above congruence and hypothesis (i) it follows that the determinant Δ_{n+1} formed for each set of $n+1$ of the m points $s_1, s_2, \cdots, s_m$ is positive. We

[12] *Loc. cit.*, footnote 7.

show that the $S_{n,r}$ does not *contain even* $n+3$ *such points,* and hence, *a fortiori,* it does not contain a set of m such points for $m > n+3$.

Denote by $\sigma^{(i)}$, $(i = 1, 2, \cdots, n+3)$, the $n+3$ vectors formed by joining the points $s_1, s_2, \cdots, s_{n+3}$ to the center of the sphere in the euclidean space E_{n+1} of $n+1$ dimensions whose surface is the $S_{n,r}$ (the sense of the vectors being taken from the center of the sphere to the points s_i). Since, as remarked above, every set of $n+1$ of the m points $s_1, s_2, \cdots, s_m$ has a positive determinant Δ_{n+1}, it follows that every set of $n+1$ of the $n+3$ vectors $\sigma^{(i)}$ is an independent set of vectors. A Cartesian coördinate system in the E_{n+1} may be introduced in terms of which the direction cosines $(\sigma_1^{(i)}, \sigma_2^{(i)}, \cdots, \sigma_{n+1}^{(i)})$ of the vectors $\sigma^{(i)}$, $(i = 1, 2, \cdots, n+3)$, are given by the following table:

$$
\text{(I)}\qquad
\begin{array}{lllllll}
\textit{vector} & & & \textit{direction cosines} & & & \\
\sigma^{(1)} & 1, & 0, & 0, & \cdot\,,\cdot\,,\cdot\,, & 0, & \\
\sigma^{(2)} & \sigma_1^{(2)}, & \sigma_2^{(2)}, & 0, & \cdot\,,\cdot\,,\cdot\,, & 0, & \sigma_2^{(2)} > 0, \\
\sigma^{(3)} & \sigma_1^{(3)}, & \sigma_2^{(3)}, & \sigma_3^{(3)}, & \cdot\,,\cdot\,,\cdot\,, & 0, & \sigma_3^{(3)} > 0, \\
\cdot & \cdot & \cdot & \cdot & \cdot\;\cdot\;\cdot & \cdot & ,\ \cdot \\
\sigma^{(n+1)} & \sigma_1^{(n+1)}, & \sigma_2^{(n+1)}, & \sigma_3^{(n+1)}, & \cdot\,,\cdot\,,\cdot\,, & \sigma_{n+1}^{(n+1)}, & \sigma_{n+1}^{(n+1)} > 0, \\
\sigma^{(n+2)} & \sigma_1^{(n+2)}, & \sigma_2^{(n+2)}, & \sigma_3^{(n+2)}, & \cdot\,,\cdot\,,\cdot\,, & \sigma_{n+1}^{(n+2)}, & \\
\sigma^{(n+3)} & \sigma_1^{(n+3)}, & \sigma_2^{(n+3)}, & \sigma_3^{(n+3)}, & \cdot\,,\cdot\,,\cdot\,, & \sigma_{n+1}^{(n+3)}, &
\end{array}
$$

with the inequalities exhibited on the right subsisting.

Since

$$\tfrac{1}{2}\pi \leqq s_i s_j / r < \pi, \qquad (i, j = 1, 2, \cdots, n+3;\ i \neq j),$$

$\cos(s_i s_j / r) = (\sigma^{(i)} \sigma^{(j)}) \leqq 0$ for each pair of distinct indices i, j, where $(\sigma^{(i)} \sigma^{(j)})$ denotes the scalar product of the vectors $\sigma^{(i)}$ and $\sigma^{(j)}$. From $(\sigma^{(1)} \sigma^{(j)}) \leqq 0$, $(j = 2, 3, \cdots, n+3)$, it is seen that $\sigma_1^{(j)} \leqq 0$, $(j = 2, 3, \cdots, n+3)$. These inequalities, together with $(\sigma^{(2)} \sigma^{(j)}) \leqq 0$, $(j = 3, 4, \cdots, n+3)$, yield $\sigma_2^{(j)} \leqq 0$, $(j = 3, 4, \cdots, n+3)$. Proceeding in this manner, it is found that all elements in the above table that lie *below* the diagonal of the first $n+1$ rows and columns are negative or zero. But

$$0 \geqq (\sigma^{(n+2)} \sigma^{(n+3)}) = \sigma_1^{(n+2)} \sigma_1^{(n+3)} + \cdots + \sigma_{n+1}^{(n+2)} \sigma_{n+1}^{(n+3)},$$

and since each summand is positive or zero, it follows that *each summand is zero.* Hence, either $\sigma_{n+1}^{(n+2)} = 0$ or $\sigma_{n+1}^{(n+3)} = 0$; that is, at least one of the vectors $\sigma^{(n+2)}, \sigma^{(n+3)}$ forms with the first n vectors $\sigma^{(1)}, \sigma^{(2)}, \cdots, \sigma^{(n)}$ a *dependent* set of $n+1$ vectors. This gives the desired contradiction.

Hence, at last one $(n+3)$-rd order principal minor of Δ_N does not vanish. Applying now Theorem 2.1, it may be concluded that every element of the

determinant outside the principal diagonal has the value $\pm 1/(n+1)$. Since $-1 < r_{ij} \leqq 0$ for each distinct pair of indices, the plus sign is excluded and the theorem is proved.

It is worth observing that for certain values of n (e. g., $n = 1, 2, 3$), the theorem just proved is valid even if the hypothesis (i) of the theorem demands merely that every principal minor of Δ_N of order less than $n+2$ be non-negative (i. e., hypothesis (i) of Theorem 2.1 may be used instead of the strengthened form adopted in Theorem 3.1). For in these cases it is possible to show (without assuming that every principal minor of order $n+1$ is positive) that at least one principal minor of order $n+3$ does not vanish. An assumption to the contrary leads, as above, to the $S_{n,r}$ containing at least $n+4$ points with pairwise distances greater than or equal to $\frac{1}{2}\pi r$ and less than πr. For $n = 1, 2, 3$, this may be shown impossible without demanding that every set of $n+1$ of the points be independent. That this circumstance is not true in general is seen by the following example of nine (unit) vectors in the euclidean E_6 that make (pairwise) angles α_{ij} such that $\frac{1}{2}\pi \leqq \alpha_{ij} < \pi$, $(i, j = 1, 2, \cdots, 9;\ i \neq j)$:

$$\sigma^{(1)} = (1, \cdot, \cdot, \cdot, \cdot, \cdot), \qquad \sigma^{(2)} = (\cdot, 1, \cdot, \cdot, \cdot, \cdot), \qquad \sigma^{(3)} = (\cdot, \cdot, 1, \cdot, \cdot, \cdot),$$

$$\sigma^{(4)} = (\cdot, \cdot, \cdot, \cdot, \cdot, 1), \qquad \sigma^{(5)} = \left(\frac{-1}{\sqrt{2}}, \cdot, \cdot, \frac{-1}{\sqrt{2}}, \cdot, \cdot\right), \quad \sigma^{(6)} = \left(\frac{-1}{\sqrt{2}}, \cdot, \cdot, \frac{1}{\sqrt{2}}, \cdot, \cdot\right),$$

$$\sigma^{(7)} = \left(\cdot, \frac{-1}{\sqrt{2}}, \cdot, \cdot, \cdot, \frac{-1}{\sqrt{2}}\right), \quad \sigma^{(8)} = \left(\cdot, \cdot, \frac{-1}{\sqrt{2}}, \cdot, \frac{-1}{\sqrt{2}}, \cdot\right), \quad \sigma^{(9)} = \left(\cdot, \cdot, \frac{-1}{\sqrt{5}}, \cdot, \frac{2}{\sqrt{5}}, \cdot\right),$$

where zero components are indicated by dots. Thus, a five dimensional spherical surface $S_{5,1}$ of unit radius contains 9 points $s_1, s_2, \cdots, s_9$ such that $\frac{1}{2}\pi \leqq s_i s_j < \pi$, $(i, j = 1, 2, \cdots, 9;\ i \neq j)$. It is unknown to the writer whether or not an analogous set of 8 points is contained in the $S_{4,r}$.

The problem suggests itself of determining the integer-valued function $f(n)$ which for each value of n gives the maximum number of points contained in the $S_{n,r}$ with pairwise distances greater than or equal to $\frac{1}{2}\pi r$ and less than πr.[13] Clearly $f(1) = 3$, $f(2) = 4$, but, as the above example shows, $f(5) \geqq 9$. It is of interest to observe that the notion of the function $f(n)$, which the metric considerations employed in the proof of Theorem 3.1 introduces quite naturally, is closely related to a concept used by A. Appert in his investigation of measure in general metric spaces.[14] This concept, which

[13] The problem can also be phrased, of course, in terms of vectors in euclidean space.

[14] A. Appert, "Mesures normales dan les espaces distanciés," *Bulletin des Sciences Mathématiques*, serie II, vol. 60 (1936), pp. 329-352; 368-380. See also "Distance geometries," pp. 135-137.

plays an important part in the theory developed, is the function $N(E,\rho)$ defined as the greatest integer n such that E (a subset of a metric space) contains n points with all $\frac{1}{2}n(n-1)$ mutual distances determined by the points exceeding the number ρ. We shall return to this matter at the end of the next Section.

4. Determinants of type Δ_{N^*}. If the elements r_{ij}, $(i \neq j)$, of a determinant of type Δ_N be restricted to lie in the interior of the interval $(-1, 0)$, the determinant is said to be of type Δ_{N^*}.

THEOREM 4. 1. *Let Δ_{N^*} be of order $m > n+3$. If* (i) *every principal minor of order less than $n+2$ is non-negative,* (ii) *every principal minor of order $n+2$ vanishes, then every element of the determinant outside the principal diagonal has the value $-1/(n+1)$.*

Proof. An application of Theorem 3. 1 gives the desired conclusion if it is shown that each principal minor of order $n+1$ is positive. We have $\Delta_{N^*} = |\cos(p_ip_j/r)|$, $(i,j = 1, 2, \cdots, m)$, with $\frac{1}{2}\pi < p_ip_j/r < \pi$ for each pair of distinct indices i, j. Suppose, now, a principal minor of order $n+1$ vanishes, and assume the labelling so that $\Delta_{n+1}(p_1, p_2, \cdots, p_{n+1}) = 0$. From the hypotheses (i), (ii), the $n+3$ points $p_1, p_2, \cdots, p_{n+3}$ are either congruent with $n+3$ points of the $S_{n,r}$, or they form a pseudo-$S_{n,r}$ set. The latter alternative is, however, impossible, for by Theorem 1. 1 the determinant Δ_{n+1} formed for any $n+1$ points of a pseudo-$S_{n,r}$ $(n+3)$-tuple does not vanish. Hence, the assumption that $\Delta_{n+1}(p_1, p_2, \cdots, p_{n+1})$ vanishes implies that the $S_{n,r}$ contains $n+3$ points $s_1, s_2, \cdots, s_{n+3}$ such that $\frac{1}{2}\pi < s_is_j/r < \pi$, $(i, j = 1, 2, \cdots, n+3;\ i \neq j)$. We assert this to be impossible.

The above assertion is evidently true for $n = 1$. If the inductive assumption of its validity is made for all positive integers $k < n$, it is shown that this implies the truth of the statement for $k = n$. Two cases present themselves.

CASE 1. *Every set of $n+1$ of the $n+3$ points $s_1, s_2, \cdots, s_{n+3}$ is a dependent set.*

Then the $n+3$ points are evidently contained in an $S_{n-1,r}$. But by the inductive hypothesis, the $S_{n-1,r}$ does not contain $n+2$ points (and hence, *a fortiori*, it does not contain $n+3$ points) with pairwise distances exceeding $\frac{1}{2}\pi r$ and less than πr.

CASE 2. *At least one set of $n+1$ of the points $s_1, s_2, \cdots, s_{n+3}$ is an independent set.*

If $s_{i_1}, s_{i_2}, \cdots, s_{i_{n+1}}$ is an independent set of $n+1$ of the points

$s_1, s_2, \cdots, s_{n+3}$, then the $n+1$ vectors $\sigma^{(i_1)}, \sigma^{(i_2)}, \cdots, \sigma^{(i_{n+1})}$, (obtained in the same manner as in Section 3) form an independent set, and a Cartesian coördinate system may be introduced in the E_{n+1} such that the table (I) of direction cosines of the $n+3$ vectors $\sigma^{(i_j)}$, $(j=1,2,\cdots,n+3)$, is obtained, with the $n+1$ independent vectors occupying the first $n+1$ rows (*with the accompanying inequalities on the quantities* $\sigma_j^{(i_j)}$, $(j=2,\cdots,n+1)$), and the remaining two vectors $\sigma^{(i_{n+2})}, \sigma^{(i_{n+3})}$ in the $(n+2)$-nd and $(n+3)$-rd rows, respectively. Since the scalar product $(\sigma^{(i_j)}\sigma^{(i_k)})$ of each vector $\sigma^{(i_j)}$ with each vector $\sigma^{(i_k)}$ following it in the table is *negative*, it is seen that each element of the table that lies below the diagonal of the first $n+1$ rows and columns is negative. But this is impossible, for then $(\sigma^{(i_{n+2})}\sigma^{(i_{n+3})}) > 0$, and the distance $s_{i_{n+2}}s_{i_{n+3}}$ does not lie between $\frac{1}{2}\pi r$ and πr, which is contrary to the condition satisfied by the points $s_1, s_2, \cdots, s_{n+3}$.

The proof of the assertion is established, then, by complete induction, and the theorem follows at once from Theorem 3. 1.

It is observed that we have shown, incidentally, that the Appert function $N(S_{n,r}, \frac{1}{2}\pi r)$ equals $n+2$. Since the argument based upon the table of direction cosines of the vectors *makes no use of the condition that* $s_i s_j < \pi r$, it is seen that the greatest number of points contained in the $S_{n,r}$ with pairwise distances exceeding $\frac{1}{2}\pi r$ is $n+2$. The remarks concluding Section 3 (and the above observation concerning the function $N(S_{n,r}, \frac{1}{2}\pi r)$) give rise quite naturally to several interesting problems.

Let $M = (s_1, s_2, \cdots, s_m)$ be a set of m points of $S_{n,r}$, and denote by $\mu(M)$ the minimum of the $\frac{1}{2}m(m-1)$ distances determined by the m points. *What is the maximum of this minimum $\mu(M)$ as M describes all subsets of m points of $S_{n,r}$?* For $m = n+2$, it is clear that

$$\max_{M \subset S_{n,r}} \min_{s_i, s_j \in M} s_i s_j = r \cdot \cos^{-1}(-1/n+1),$$

the length of the "side" of the equilateral $(n+2)$-tuple contained in the $S_{n,r}$. A similar remark may be made whenever $m \leqq n+1$, by considering an equilateral set of m points on a lower-dimensional "great sphere" $S_{k,r}$ of $S_{n,r}$. For $m > n+2$, n an arbitrary integer, the above question does not seem to be easily answered.

For $n=2$, the problem may be expressed in terms of a number p of planes on a point and the $3\binom{p}{3}$ angles formed by them. This phraseology admits the following plane dual:

What is the minimum of the maximum of the $3\binom{p}{3}$ angles arising from a set M of p points of a plane when M describes all planar subsets of p points?

For $p = 3, 4, 5, 6$, the answer to this query is given by the function $(p - 2)\pi/p$, but for $p = 7$ *this formula fails to give the minimum maximum angle.* There exist configurations of 7 points in the plane with the largest angle arbitrarily close to $2\pi/3$.[15]

5. Determinants of type $\Delta_P = | r_{ij} |$, $r_{ij} = r_{ji}$, $0 \leqq r_{ij} < 1$, $(i \neq j)$, $r_{ii} = 1$, $(i, j = 1, 2, \cdots, m)$.

THEOREM 5.1. *Let* Δ_P *be of order* $m > n + 3$. *If* (i) *every principal minor of order less than* $n + 2$ *is non-negative,* (ii) *every principal minor of order* $n + 2$ *vanishes, then the rank of* Δ_P *does not exceed* $n + 1$.

Proof. It suffices to show that every $(n + 3)$-rd order principal minor of the determinant vanishes. Writing $\Delta_P = \Delta_m(p_1, p_2, \cdots, p_m)$, assume that there exists a non-vanishing principal minor of order $n + 3$. Then the m points $p_1, p_2, \cdots, p_m$ form a pseudo-$S_{n,r}$ set of $m > n + 3$ points, no two of which are diametral, and hence, according to Theorem 1.6, every element of Δ_P outside the principal diagonal has the value $\pm 1/(n + 1)$. Since $0 \leqq r_{ij} < 1$, $(i, j = 1, 2, \cdots, m;\ i \neq j)$, it follows that $r_{ij} = 1/(n + 1)$ for every pair of distinct indices i, j. But this is impossible, for then each $(n + 2)$-nd order principal minor of Δ_P has the value $2[n/(n + 1)]^{n+1}$, and consequently does not vanish. This contradicts (ii), and the theorem is proved.

It would be interesting to obtain purely algebraic proofs for the determinant theorems proved in this note by geometric methods. This is particularly desirable for Theorem 2.1, which arises from the distance geometry problem of characterizing metrically pseudo-$S_{n,r}$ sets of points, and which furnishes the key to the other theorems treated in this paper.

THE UNIVERSITY OF MISSOURI.

---

[15] This interesting fact was communicated to the writer by P. Erdös.

ON THE IRREDUCIBILITY OF CERTAIN CLASSES OF POLYNOMIALS.* †

By Benjamin Rosenbaum.

I. Introduction.

The following classes of polynomials were shown by I. Schur [1] to be irreducible in the rational field:

$$(1) \qquad f_1(x) = 1 + g_1 \frac{x}{1!} + g_2 \frac{x^2}{2!} + \cdots + g_{n-1} \frac{x^{n-1}}{(n-1)!} \pm \frac{x^n}{n!},$$

$$(2) \qquad f_2(x) = 1 + g_1 \frac{x^2}{u_2} + g_2 \frac{x^4}{u_4} + \cdots + g_{n-1} \frac{x^{2n-2}}{u_{2n-2}} \pm \frac{x^{2n}}{u_{2n}},$$

where the g_ν are arbitrary, rational integers and $u_{2\nu} = 1 \cdot 3 \cdot 5 \cdot \cdots \cdot (2\nu - 1)$. In (2) n is greater than 1.

Examples of (1) are the expansions of e^x, $\cos x$, and $1 \pm \sin x$, limited to a finite number of terms, as well as the polynomials of Laguerre defined by

$$L(x) = \frac{e^x}{n!} \frac{d^n(x^n e^{-x})}{dx^n} = \sum_{\nu=0}^{n} (-1)^\nu \binom{n}{\nu} \frac{x^\nu}{\nu!}.$$

Examples of (2) are the Hermitian polynomials defined by

$$H_m(x) = (-1)^m e^{x^2/2} \frac{d^m(e^{-x^2/2})}{dx^m} = \sum_{\nu=0}^{[m/2]} (-1)^\nu \binom{m}{2\nu} u_{2\nu} x^{m-2\nu},$$

for even degree $m > 2$, after multiplication by $(-1)^{m/2}/u_m$, where $[m/2]$ represents the largest integer $\leq m/2$.

Schur [2] also proved the irreducibility (with certain exceptions) of the polynomials:

$$(3) \qquad f_3(x) = 1 + g_1 \frac{x}{2!} + g_2 \frac{x^2}{3!} + \cdots + g_{n-1} \frac{x^{n-1}}{n!} \pm \frac{x^n}{(n+1)!},$$

$$(4) \qquad f_4(x) = 1 + g_1 \frac{x^2}{u_4} + g_2 \frac{x^4}{u_6} + \cdots + g_{n-1} \frac{x^{2n-2}}{u_{2n}} \pm \frac{x^{2n}}{u_{2n+2}}.$$

* Received September 1, 1938.

† Presented to the American Mathematical Society, October 26, 1935.

[1] I. Schur, "Einige Sätze über Primzahlen mit Anwendungen auf Irreduzibilitätsfragen," I, II, *Sitzungsberichte der preuss. Akad. d. Wissensch., phys.-math. Klasse* (1929), pp. 125-136, 370-391. We shall refer to these papers as Schur I and Schur II, respectively.

[2] Schur II.

Examples of irreducible polynomials of type (3) are the expansions of $(\sin x)/x$ and $(e^x - 1)/x$, limited to a finite number of terms.

Examples of irreducible polynomials of type (4) are the Hermitian polynomials $H_m(x)$ for odd m, after multiplication by $\dfrac{(-1)^{(m-1)/2}}{u_{m+1}x}$.

In the present paper we shall generalize polynomials (1) and (2) and prove the following theorems:

THEOREM 1. *Every polynomial of the form*

$$(5) \quad f(x) = \frac{g_0}{d_0} + g_1 \frac{x^r}{d_1(s-t)\,!} + g_2 \frac{x^{2r}}{d_2(2s-t)\,!} + \cdots + g_n \frac{x^{nr}}{d_n(ns-t)\,!},$$

where d_ν, g_ν, n, r, s, and t are rational integers with certain restrictions, is irreducible in the rational field. The restrictions are: n, r, and s are arbitrary, positive integers. n is positive and ≥ 2 when $r \geq 2$. $0 \leq t \leq ns - 2$. (When $\nu s - t < 1$, $(\nu s - t)\,! = 1$.) The g_ν, $1 \leq \nu \leq n-1$, are arbitrary integers. g_0 and g_n are divisible only by primes $> ns - t$ or $<$ the largest prime dividing $ns - t$. d_0 is any factor of $(p-1)\,!$, where p is the largest prime $\leq ns - t$. d_ν is the least positive integer such that $d_\nu(\nu s - t)\,!$ is divisible by d_0.

Examples of irreducible polynomials of type (5) are the expansions, limited to a finite number of terms, of e^{x^r}, $\cos x^r$, $g_0/d_0 \pm x^{rt} \sin x^r$, $g_0/d_0 \pm x^{rt} \cos x^r$ and $g_0/d_0 \pm x^{rt} e^{x^r}$, as well as $g_0/d_0 \pm x^t L$, where r and t are any positive integers and L is a polynomial of Laguerre.

THEOREM 2. *Every polynomial of the form*

$$(6) \quad g(x) = \frac{g_0}{d_0} + g_1 \frac{x^r}{d_1 u_{2(s-t)}} + g_2 \frac{x^{2r}}{d_2 u_{2(2s-t)}} + \cdots + g_n \frac{x^{nr}}{d_n u_{2(ns-t)}},$$

where d_ν, g_ν, n, r, s, and t are rational integers with certain restrictions, is irreducible in the rational field. The restrictions are: n, r and s are arbitrary, positive integers. n is positive and ≥ 2 when $r \geq 2$. $0 \leq t \leq ns - 2$. (When $\nu s - t < 1$, $u_{2(\nu s-t)} = 1$. The g_ν, $1 \leq \nu \leq n-1$, are arbitrary integers. g_0 and g_n are divisible only by primes $> 2(ns-t) - 1$ or $<$ the largest prime dividing $2(ns-t)-1$. d_0 is any factor of $1 \cdot 3 \cdot 5 \cdot \cdots \cdot (p-2)$, where p is the largest prime $\leq 2(ns-t)-1$. d_ν is the least positive integer such that $d_\nu u_{2(\nu s-t)}$ is divisible by d_0.

Remark. It is obvious that $d_n = 1$ in polynomials (5) and (6).

Examples. Let

$$f_{2n} = \frac{(-1)^n H_{2n}(x)}{u_{2n}} = \sum_{\nu=0}^{n} (-1)^\nu \binom{n}{\nu} \frac{x^{2\nu}}{u_{2\nu}}$$

and

$$g_{2n} = \frac{(-1)^n H_{2n+1}(x)}{u_{2n+2}x} = \sum_{\nu=0}^{n} (-1)^\nu \binom{n}{\nu} \frac{x^{2\nu}}{u_{2\nu+2}},$$

where $H_m(x)$ represents a polynomial of Hermite. Then $g_0/d_0 \pm x^{2t}f_{2n}$ and $g_0/d_0 \pm x^{2t}g_{2n}$ are irreducible polynomials of class (6).

We shall also discuss the application of the Schoenemann [3]-Eisenstein [4] and Koenigsberger [5] irreducibility criteria to certain cases of polynomials (5) and (6).

II. Generalizations of polynomial (1).

1. Preliminary theorems and corollary. The theorem which follows was first stated and proved by Sylvester.[6] Simpler proofs have been given by I. Schur [7] and P. Erdös.[8]

THEOREM 3. *When $k \geq 1$, in every set of k consecutive integers $\geq k+1$, there is at least one integer divisible by a prime $\geq k+1$.*

COROLLARY 1. *When $k \geq 2$, in every set of k consecutive integers $\geq k$, there is at least one integer divisible by a prime $\geq k+1$.*

Proof. We need only consider the case where the smallest integer of the sequence is k. By the theorem, one of the last k of the $k+1$ consecutive integers $k, k+1, \cdots, 2k-1, 2k$, is divisible by a prime $\geq k+1$. Since $2k$ is not divisible by this prime, it must divide one of the first k integers, which proves the corollary.

THEOREM [9] 4. *If h_n denotes the index of the highest power of the prime p dividing $n!$ and σ denotes the sum of the digits of n to the base p, then $h_n = (n-\sigma)/(p-1)$.*

2. First generalization of polynomial (1). We shall prove

THEOREM 5. *Every polynomial of the form*

---

[3] T. Schoenemann, "Zur Theorie der höhern Congruenzen," *Journal für Mathematik*, vol. 32 (1846), p. 100.

[4] G. Eisenstein, "Über die Irreductibilität und einige andere Eigenschaften der Gleichung, von welcher die Theilung der ganzen Lemniscate abhängt," *Journal für Mathematik*, vol. 39 (1850), pp. 166, 167.

[5] L. Koenigsberger, "Über den Eisensteinschen Satz von der Irreductibilität algebraischer Gleichungen," *Journal für Mathematik*, vol. 115 (1895), pp. 53-78.

[6] J. J. Sylvester, "On arithmetical series," *Messenger of Mathematics*, vol. 21 (1892), pp. 1-19, 87-120; *Collected Mathematical Papers*, vol. 4, pp. 687-731.

[7] Schur I, pp. 128-136.

[8] P. Erdös, "A theorem of Sylvester and Schur," *Journal of the London Mathematical Society*, vol. 9 (1934), pp. 282-288.

[9] A. M. Legendre, *Théorie des nombres*, Ed. 2 (1808), p. 8.

$$(7) \qquad f(x) = \sum_{\nu=0}^{n-1} g_\nu \frac{x^{\nu r}}{d_\nu(\nu - t)\,!} \pm \frac{x^{nr}}{d_n(n - t)\,!},$$

with the restrictions given in Theorem 1, is irreducible in the rational field.

Proof. In considering the question of irreducibility we may replace (7) by a polynomial with integral coefficients, that of the highest power of x being unity:

$$(8) \qquad F(x) = \pm (n-t)\,!\, f(x) = x^{nr} \pm \sum_{\nu=0}^{n-1} g_\nu \frac{(n-t)\,!\, x^{\nu r}}{d_\nu(\nu - t)\,!}.$$

If $F(x)$ is reducible, it has an irreducible factor

$$A(x) = x^k + a_1 x^{k-1} + \cdots + a_{k-1}x + a_k$$

with rational, integral coefficients and degree $k \leq nr/2$.

We shall now proceed with some lemmas:

LEMMA [10] 1. *If α is a root of $A(x) = 0$ and p is a rational prime dividing a_k, then the principal ideals $[\alpha]$ and $[p]$ are not relatively prime in the field $K(\alpha)$ of degree k.*

LEMMA 2. $\qquad a_k \equiv 0 \pmod{n-t}.$

Proof. $F(x) = A(x)B(x) \equiv x^{nr} \pmod{n-t}$. Since the degree of $A(x) = k$, $A(x) \equiv x^k \pmod{n-t}$, and $a_k \equiv 0 \pmod{n-t}$.

LEMMA 3. *If p is the largest prime dividing $(n-t)\,!$ and a_k, then $p \geqq 2$ and does not divide g_0.*

Proof. Since $n-t$ divides $(n-t)\,!$ and a_k, $p \geq$ the largest prime dividing $n-t$. The lemma follows from the restrictions of Theorem 1: 1) $n - t \geq 2$; 2) g_0 is divisible only by primes $> n-t$ or $<$ the largest prime dividing $n-t$.

LEMMA 4. *If p is any prime dividing $(n-t)\,!$ and a_k, and i is an integer such that $kt \geq n(i-1)$, then $p \leq (k+r-i)/r$.*

Proof. Take p as the largest prime dividing $(n-t)\,!$ and a_k. Let $\mathfrak{p}$ be a prime ideal dividing the ideals $[\alpha]$ and $[p]$. If $\mathfrak{p}^e$ and $\mathfrak{p}^f$ are the highest powers of $\mathfrak{p}$ dividing $[\alpha]$ and $[p]$ respectively, then $e \geq 1$ and $1 \leq f \leq k$, since the norm of p in the field $K(\alpha) = p^k$. Now consider

$$F(\alpha) = \alpha^{nr} \pm \sum_{\nu=0}^{n-1} g_\nu \frac{(n-t)\,!\, \alpha^{\nu r}}{d_\nu(\nu - t)\,!} = 0.$$

Since $F(\alpha) \equiv 0 \pmod{\mathfrak{p}^\delta}$ for any δ, the highest power $(\mathfrak{p}^\epsilon)$ of $\mathfrak{p}$ which divides $g_0(n-t)\,!/d_0 \geq$ the highest power of $\mathfrak{p}$ which divides some other term

[10] Schur I, p. 126.

$g_\nu \dfrac{(n-t)!\,\alpha^{\nu r}}{d_\nu(\nu-t)!}$, $1 \leq \nu \leq n$. Otherwise $F(\alpha) \not\equiv 0 \pmod{\mathfrak{p}^{e+1}}$. Since p does not divide g_0, and d_ν is a factor of d_0, the power of $\mathfrak{p}$ dividing $(n-t)! \geq$ the power of $\mathfrak{p}$ dividing $(n-t)!\,\alpha^{\nu r}/(\nu-t)!$. Now $(n-t)! = p^{h_{n-t}}q = \mathfrak{p}^{f(h_{n-t})}Q$ and $(n-t)!\,\alpha^{\nu r}/(\nu-t)! = \mathfrak{p}^{f(h_{n-t}-h_{\nu-t})+er\nu}R$, where $q \not\equiv 0 \pmod p$ and Q and R are ideals $\not\equiv 0 \pmod{\mathfrak{p}}$. Hence $f(h_{n-t}-h_{\nu-t}) + er\nu \leq fh_{n-t}$ and $\nu r \leq kh_{\nu-t}$. By Theorem 4, $h_{\nu-t} < (\nu-t)/(p-1)$. Therefore

$$\nu r < k(\nu-t)/(p-1)$$

and

$$p < \frac{\nu(k+r)}{\nu r} - \frac{kt}{\nu r} \leq \frac{k+r}{r} - \frac{kt}{nr} \leq \frac{k+r}{r} - \frac{n(i-1)}{nr} = \frac{k+r-(i-1)}{r}.$$

COROLLARY 2. *Set $k = \kappa r + \epsilon_1$, $0 \leq \epsilon_1 < r$, and $i = \mu r + \epsilon_2$, $0 \leq \epsilon_2 < r$. Then $p \leq \kappa - \mu + 1$ when $\epsilon_1 \geq \epsilon_2$, and $p \leq \kappa - \mu$ when $\epsilon_1 < \epsilon_2$. Also $\kappa \geq \mu + 1$ when $\epsilon_1 \geq \epsilon_2$ and $\kappa \geq \mu + 2$ when $\epsilon_1 < \epsilon_2$.*

Proof. To prove the last part of the corollary we make use of the fact that $p \geq 2$.

LEMMA 5. *If i is an integer such that $kt < ni$, then $n - t > 2(k-i)/r$.*

Proof. When $i > rt/2$, since $k \leq nr/2$, $k - i \leq (nr - 2i)/2 < (nr - rt)/2$ and $n - t > 2(k-i)/r$. When $i \leq rt/2$, since $kt < ni$, $t(k-i) < (n-t)i \leq (n-t)rt/2$ and again $n - t > 2(k-i)/r$.

COROLLARY 3. *Set $k = \kappa r + \epsilon_1$, $0 \leq \epsilon_1 < r$, and $i = \mu r + \epsilon_2$, $0 \leq \epsilon_2 < r$. Then $n - t > 2\kappa - 2\mu + 2(\epsilon_1 - \epsilon_2)/r$.*

LEMMA 6. *Let p be the largest prime dividing $(n-t)(n-t-1)\cdot\cdots\cdot(n-t-j+1)$. Set $k = \kappa r + \epsilon_1$, $0 \leq \epsilon_1 < r$. When $\epsilon_1 = 0$, take $j \leq \kappa$; when $\epsilon_1 \geq 1$, take $j \leq \kappa + 1$. Then $a_k \equiv 0 \pmod p$.*

Proof. Consider

$$\begin{aligned} F(x) = x^{nr} \pm (&g_{n-1}(n-t)x^{(n-1)r}/d_{n-1} + g_{n-2}(n-t)(n-t-1)x^{(n-2)r}/d_{n-2} \\ &+ \cdots + g_{n-j}(n-t)(n-t-1)\cdot\cdots\cdot(n-t-j+1)x^{(n-j)r}/d_{n-j} \\ &+ \cdots + g_0(n-t)!/d_0). \end{aligned}$$

Let q be the largest prime $\leq n - t$.

When $n - t - j \geq q - 1$, $(n-t-j)! \equiv 0 \pmod{d_0}$ and $d_{n-j} = 1$. (See restrictions of Theorem 1.) Hence p divides

$$(n-t)(n-t-1)\cdots(n-t-j+1)/d_{n-j}.$$

When $n - t - j < q - 1$, $(n-t)(n-t-1)\cdot\cdots\cdot(n-t-j+1)$

is divisible by q. Hence $p = q$. d_{n-j}, being a factor of $(q-1)!$, is not divisible by p. Hence again p divides

$$(n-t)(n-t-1)\cdot\cdots\cdot(n-t-j+1)/d_{n-j}.$$

Since $d_{n-l}(n-t-l)!$, $l > j$, is a factor of $d_{n-j}(n-t-j)!$, p also divides

$$(n-t)!/(d_{n-l}(n-t-l)!) = (n-t)(n-t-1)\cdot\cdots\cdot(n-t-l+1)/d_{n-l}.$$

Hence $F(x) \equiv x^{(n-j+1)r}G(x) \pmod{p}$. $F(x) = A(x)B(x)$, the degree of $B(x)$ being equal to $nr-k$. Thus, if $(n-j+1)r > nr-k$, $A(x) \equiv xH(x) \pmod{p}$ and $a_k \equiv 0 \pmod{p}$. Since $(n-j+1)r > nr-k$ when $j < (r+k)/r = \kappa + 1 + \epsilon_1/r$, the lemma follows.

Contradiction. In the product of Lemma 6 take $j = \kappa$ when $\epsilon_1 = 0$ and $j = \kappa + 1$ when $\epsilon_1 \geq 1$. Then $a_k \equiv 0 \pmod{p}$.

Let i be an integer such that $n(i-1) \leq kt < ni$. Then Lemmas 4 and 5 and their corollaries apply. By Corollary 3, $n-t-j+1 > \kappa - 2\mu + 1 + 2(\epsilon_1-\epsilon_2)/r$ when $j = \kappa$, and $> \kappa - 2\mu + 2(\epsilon_1 - \epsilon_2)/r$ when $j = \kappa + 1$. In either case the product of Lemma 6 contains κ consecutive integers $> \kappa - 2\mu + 1 + 2(\epsilon_1 - \epsilon_2)/r$.

When $\epsilon_1 \geq \epsilon_2$ the product contains κ consecutive integers $\geq \kappa - 2\mu + 2$, or $\kappa - \mu + 1$ consecutive integers $\geq \kappa - \mu + 1$. By Corollary 2, $\kappa - \mu + 1 \geq 2$. Therefore, by Corollary 1, $p \geq \kappa - \mu + 2$. But, by Corollary 2, $p \leq \kappa - \mu + 1$. Thus we have a contradiction arising from the assumption that polynomial (8) is reducible. Consequently (8) and (7) are irreducible.

When $\mu = 0$, $\kappa - \mu + 1 = \kappa + 1$ and $j = \kappa + 1$. If, in addition, $\epsilon_1 = 0$, a_k is not necessarily $\equiv 0 \pmod{p}$. But when $\mu = 0$, $\epsilon_2 \geq 1$, otherwise $i = \mu r + \epsilon_2 = 0$ and $kt < ni = 0$. Hence $\epsilon_1 < \epsilon_2$ and the next paragraph applies.

When $\epsilon_1 < \epsilon_2$, $n - t - \kappa + 1 \geq \kappa - 2\mu$. Hence we have κ consecutive integers $\geq \kappa - 2\mu$ or $\kappa - \mu$ consecutive integers $\geq \kappa - \mu$. By Corollary 2, $\kappa - \mu \geq 2$. Therefore, by Corollary 1, $p \geq \kappa - \mu + 1$. But, by Corollary 2, $p \leq \kappa - \mu$. Thus we again have a contradiction and polynomials (8) and (7) are irreducible, which proves Theorem 5.

3. Further generalizations of polynomial (1).

We now prove

THEOREM 6. *Every polynomial of the form*

$$f(x) = \sum_{\nu=0}^{n} g_\nu \frac{x^{\nu r}}{d_\nu(\nu - t)!}, \tag{9}$$

with the restrictions given in Theorem 1, is irreducible in the rational field.

Remark. The difference between polynomials (7) and (9) lies in the coefficient g_n, which in the case of (9) may not equal unity.

Proof. If we multiply (9) by g_n^{nr-1} and replace $g_n x$ by x, we obtain a polynomial of type (7), which is irreducible. The same is therefore true of (9).

As a final generalization we have Theorem 1, stated in the Introduction.

Proof of Theorem 1. Consider the following irreducible ploynomial of type (9) and degree nrs, with $g_\nu = 0$ when $\nu \neq ms$:

$$f(x) = \frac{g_0}{d_0} + g_s \frac{x^{rs}}{d_s(s-t)\,!} + g_{2s} \frac{x^{2rs}}{d_{2s}(2s-t)\,!} + \cdots + g_{ns} \frac{x^{nrs}}{d_{ns}(ns-t)\,!}.$$

Substituting x for x^s and g_m for g_{ms}, we obtain a polynomial of type (5), which is therefore irreducible.

4. Application of the Koenigsberger criterion to polynomial (5).

KOENIGSBERGER THEOREM. *The polynomial*

$$f(x) = g_0 x^n + g_1 p^{[e/n]+1} x^{n-1} + g_2 p^{[2e/n]+1} x^{n-2} \\ + \cdots + g_\nu p^{[\nu e/n]+1} x^{n-\nu} + \cdots + g_n p^e,$$

where the g_ν are rational integers, g_0 and g_n not divisible by p, and e relatively prime to n, is irreducible in the rational field.

Under certain conditions we can use the above theorem to establish the irreducibility of polynomials of type (5), with less restriction on g_0 and g_n than given in Theorem 1. We shall prove

THEOREM 7. *Any polynomial $f(x)$ of type (5), with $ns-t$ a power of a prime p; g_0 and g_n relatively prime to p; e, the index of the highest power of p dividing $(ns-t)\,!/d_0$, relatively prime to nr; and the remaining restrictions of Theorem 1 applying, is irreducible in the rational field.*

Proof. Set g_ν of $f(x)$ *equal to $g_{n-\nu}$ and consider*

$$F(x) = (ns-t)\,!\,f(x) = \sum_{\nu=0}^{n} g_\nu \frac{(ns-t)\,!\,x^{(n-\nu)r}}{d_{n-\nu}((n-\nu)s-t)\,!}.$$

Denote the indices of the highest powers of p dividing $d_{n-\nu}$ and the coefficient of $x^{(n-\nu)r}$ by $i_{n-\nu}$ and $j_{n-\nu}$ respectively.

By Theorem 4, $e = (ns-t-1)/(p-1) - i_0$, and

$$j_{n-\nu} \geq (ns-t-1)/(p-1) - (s(n-\nu)-t-\sigma)/(p-1) - i_{n-\nu},$$

where σ denotes the sum of the digits of $s(n-\nu)-t$ to the base p.

By the definition of $d_{n-\nu}$, (see Theorem 1), $(s(n-\nu)-t-\sigma)/(p-1)+i_{n-\nu}=i_0$ when $(s(n-\nu)-t-\sigma)/(p-1)<i_0$, and $i_{n-\nu}=0$ when $(s(n-\nu)-t-\sigma)/(p-1)\geq i_0$.

In the first case $j_{n-\nu}\geq e\geq[\nu e/n]+1$, where $1\leq\nu\leq n-1$.

In the second case

$$j_{n-\nu}\geq(\nu s+\sigma-1)/(p-1)\geq[(\nu s-1)/(p-1)]+1,$$

since $(\nu s+\sigma-1)/(p-1)$ is an integer. Also

$$\begin{aligned}[\nu e/n]&\leq[(\nu/n)(ns-1)/(p-1)]\\&=[(\nu s-\nu/n)/(p-1)]=[(\nu s-1)/(p-1)].\end{aligned}$$

Hence again $j_{n-\nu}\geq[\nu e/n]+1$.

The conditions of the Koenigsberger theorem being fulfilled, $F(x)$ is irreducible.

Corollary 4. *Sufficient conditions for e to be relatively prime to nr are: 1) d_0 relatively prime to p; 2) $ns-t$ and nr powers of p.*

Proof.

$$e=(ns-t-1)/(p-1)=\frac{p^a-1}{p-1}=p^{a-1}+p^{a-2}+\cdots+p+1.$$

Hence e is not divisible by p and is relatively prime to nr.

Corollary 5. *Every polynomial of the form*

$$f(x)=g_0+g_1\frac{x}{1!}+g_2\frac{x^2}{2!}+\cdots+g_n\frac{x^n}{n!},$$

where n is the power of a prime p, g_0 and g_n relatively prime to p, and the remaining coefficients arbitrary integers, is irreducible in the rational field.

III. Generalizations of polynomial (2).

5. Preliminary theorems and corollaries. We shall make use of the following:

Theorem [11] 8. *When $k\geq 3$, in every set of k consecutive, odd integers $\geq 2k+3$, there is at least one integer divisible by a prime $\geq 2k+3$. When $k=2$, the sole exception is the pair of integers 25, 27. When $k=1$, the only exceptions are integers of the form 3^α, with $\alpha\geq 2$.*

Corollary 6. *When $k\geq 1$, in every set of $k+1$ consecutive, odd integers $\geq 2k+1$, there is at least one integer divisible by a prime $\geq 2k+3$.*

[11] Schur II, pp. 372-379.

Proof. Since there are $k+1$ consecutive, odd integers $\geq 2k+1$, they must include k consecutive, odd integers $\geq 2k+3$. The corollary then follows from the theorem, with the exceptions noted for $k=1$ and $k=2$.

When $k=2$, consider the pair of integers 25, 27. In this case the $k+1$ integers of the corollary are 23, 25, 27. The corollary is true since $23 > 2k+3 = 7$.

When $k=1$, consider integers of the form 3^α, $\alpha \geq 2$. The $k+1$ integers of the corollary are $3^\alpha - 2, 3^\alpha$. The corollary again holds since $3^\alpha - 2$ is divisible only by primes $\geq 5 = 2k+3$.

COROLLARY 7. *When $k \geq 2$, in every set of k consecutive, odd integers $\geq 2k-1$, there is at least one integer divisible by a prime $\geq 2k+1$.*

This is a restatement of Corollary 6 with k substituted for $k+1$.

THEOREM [12] 9. *If λ_n denotes the index of the highest power of the prime p dividing $u_{2n} = 1 \cdot 3 \cdot 5 \cdot \cdots \cdot (2n-1)$, then $\lambda_n < 2n/p$.*

6. First generalization of polynomial (2). We shall prove

THEOREM 10. *Every polynomial of the form*

$$g(x) = \sum_{\nu=0}^{n-1} g_\nu \frac{x^{\nu r}}{d_\nu u_{2(\nu-t)}} \pm \frac{x^{nr}}{d_n u_{2(n-t)}}, \tag{10}$$

with the restrictions given in Theorem 2, is irreducible in the rational field.

Proof. As in § 2 we shall deal with an integral polynomial of which the coefficient of the highest power of x is unity:

$$G(x) = \pm u_{2(n-t)} g(x) = x^{nr} \pm \sum_{\nu=0}^{n-1} g_\nu \frac{u_{2(n-t)} x^{\nu r}}{d_\nu u_{2(\nu-t)}}, \tag{11}$$

and consider an irreducible factor $A(x)$, with constant term a_k and degree $k \leq nr/2$. Then the methods of § 2 can be used to prove the following lemmas. (Most of the proofs have been omitted or shortened since they can be easily reproduced from the corresponding proofs of § 2.)

LEMMA 7. $\quad a_k \equiv 0 \pmod{2(n-t)-1}$.

LEMMA 8. *If p is the largest prime dividing $u_{2(n-t)}$ and a_k, then p does not divide g_0. Also $p \geq 3$.*

LEMMA 9. *If p is any prime dividing $u_{2(n-t)}$ and a_k, and i is an integer such that $2kt \geq n(i-1)$, then $p \leq (2k-i)/r$.*

Proof. We proceed as in the proof of Lemma 4 and find $\nu r \leq k\lambda_{\nu-t}$. Hence, by Theorem 9, $\nu r < 2k(\nu - t)/p$, from which we find

[12] Schur II, pp. 380, 381.

$$p < 2k/r - 2kt/\nu r \leq 2k/r - 2kt/nr \leq (2k - (i-1))/r.$$

COROLLARY 8. *Let* $k = \kappa r + \epsilon_1$, $0 \leq \epsilon_1 < r$, *and* $i = 2\mu r + \epsilon_2$, $0 \leq \epsilon_2 < 2r$. *Then* $p \leq 2\kappa - 2\mu + 1$ *and* $\kappa \geq \mu + 1$ *when* $2\epsilon_1 \geq \epsilon_2$; $p \leq 2\kappa - 2\mu - 1$ *and* $\kappa \geq \mu + 2$ *when* $2\epsilon_1 < \epsilon_2$.

Proof. The relations between κ and μ are derived by making use of $p \geq 3$.

LEMMA 10. *If* i *is an integer such that* $2kt < ni$, *then* $n - t > (2k - i)/r$.

Proof. When $i > rt$, since $k \leq nr/2$, $n - t > 2k/r - i/r = (2k - i)/r$. When $i \leq rt$, since $2kt < ni$,

$$n - t > 2kt/i - t \geq (i/r)((2k - i)/i) = (2k - i)/r.$$

COROLLARY 9. *Let* $k = \kappa r + \epsilon_1$, $0 \leq \epsilon_1 < r$, *and* $i = 2\mu r + \epsilon_2$, $0 \leq \epsilon_2 < 2r$. *When* $2kt < ni$, $n - t > 2\kappa - 2\mu + (2\epsilon_1 - \epsilon_2)/r$.

LEMMA 11. *Let* $p =$ *the largest prime dividing the product of* j *factors:*

$$(2(n-t) - 1)(2(n-t) - 3) \cdot \cdots \cdot (2(n-t) - 2j + 1).$$

Set $k = \kappa r + \epsilon_1$, $0 \leq \epsilon_1 < r$. *When* $\epsilon_1 = 0$, *take* $j \leq \kappa$; *when* $\epsilon_1 \geq 1$, *take* $j \leq \kappa + 1$. *Then* $a_k \equiv 0 \pmod{p}$.

Contradiction. In the product of Lemma 11, take $j = \kappa$ when $\epsilon_1 = 0$ and $j = \kappa + 1$ when $\epsilon_1 \geq 1$. Then $a_k \equiv 0 \pmod{p}$.

Let i be an integer such that $n(i - 1) \leq 2kt < ni$. Then Lemmas 9 and 10 and their corollaries apply. By Corollary 9,

$$2(n-t) - 2j + 1 > 2\kappa - 4\mu + 1 + 2(2\epsilon_1 - \epsilon_2)/r$$

when $j = \kappa$, and $> 2\kappa - 4\mu - 1 + 2(2\epsilon_1 - \epsilon_2)/r$ when $j = \kappa + 1$. In either case the product contains κ consecutive, odd integers $> 2\kappa - 4\mu + 1 + 2(2\epsilon_1 - \epsilon_2)/r$.

When $2\epsilon_1 \geq \epsilon_2$ we have κ consecutive, odd integers $\geq 2\kappa - 4\mu + 3$ or $\kappa - \mu + 1$ consecutive, odd integers $\geq 2\kappa - 2\mu + 1$. By Corollary 8, $\kappa - \mu + 1 \geq 2$. Hence by Corollary 7, $p \geq 2\kappa - 2\mu + 3$. But, by Corollary 8, $p \leq 2\kappa - 2\mu + 1$. Thus we have a contradiction, due to the assumption that polynomial (11) is reducible. Therefore (11) and consequently (10) is irreducible.

When $\mu = 0$, $\kappa - \mu + 1 = \kappa + 1$ and $j = \kappa + 1$. If, in addition, $\epsilon_1 = 0$, a_k is not necessarily $\equiv 0 \pmod{p}$. (See Lemma 11.) But when $\mu = 0$, $\epsilon_2 > 0$, otherwise $i = 2\mu r + \epsilon_2 = 0$ and $2kt < ni = 0$. Hence when $\epsilon_1 = 0$, $2\epsilon_1 < \epsilon_2$, which case is dealt with in the next paragraph.

When $2\epsilon_1 < \epsilon_2$,

$$2(n-t)-2\kappa+1>2\kappa-4\mu-3,$$

since $2(\epsilon_2-2\epsilon_1)/r<4$. Hence we have κ consecutive, odd integers $\geq 2\kappa-4\mu-1$ or $\kappa-\mu$ consecutive, odd integers $\geq 2\kappa-2\mu-1$. By Corollary 8, $\kappa-\mu\geq 2$. Therefore, by Corollary 7, $p\geq 2\kappa-2\mu+1$. But, by Corollary 8, $p\leq 2\kappa-2\mu-1$. Thus we again have a contradiction and polynomials (11) and (10) are irreducible, which proves Theorem 10.

7. Further generalizations of polynomial (2). We now prove

THEOREM 11. *Every polynomial of the form*

$$g(x)=\sum_{\nu=0}^{n} g_\nu \frac{x^{\nu r}}{d_\nu u_{2(\nu-t)}}, \tag{12}$$

with the restrictions given in Theorem 2, is irreducible in the rational field.

Remark. Polynomial (12) differs from polynomial (10) in the coefficient g_n, which in the case of (12) may not equal unity.

Proof. The proof of Theorem 11 is similar to that of Theorem 6.

As a final generalization we have Theorem 2, stated in the Introduction.

Proof of Theorem 2. This theorem follows from Theorem 11 in the same way that Theorem 1 follows from Theorem 6.

8. Application of other irreducibility criteria to polynomial (6).

SCHOENEMANN-EISENSTEIN CRITERION. When $2(ns-t)-1$ is a prime p, each coefficient of $G(x)=u_{2(ns-t)}g(x)$, other than g_n, is divisible by p, while the constant term $u_{2(ns-t)}g_0/d_0$ is not divisible by p^2. Hence $G(x)$ and consequently $g(x)$ are irreducible.

KOENIGSBERGER CRITERION. The application of this criterion to polynomial (6), when $2(ns-t)-1$ or $ns-t$ are powers of a prime p higher than the first power, does not yield a theorem analogous to Theorem 7. This is illustrated by the following

Example. Consider $G(x)=u_{50}g(x)=u_{50}\sum_{\nu=0}^{25} x^\nu/u_{2\nu}$. In this case $ns-t=n=25$ and $2(ns-t)-1=2n-1=49=7^2$. The only prime dividing the coefficient of x^{24} is 7. The highest power of 7 dividing the coefficient of x^5 is 7^4. If $F(x)$ were a Koenigsberger polynomial this coefficient would be divisible by 7^5.

YALE UNIVERSITY.

THE FOURIER-STIELTJES COEFFICIENTS OF A FUNCTION OF BOUNDED VARIATION.*

By A. C. Schaeffer.

Let $F(x)$ be a continuous monotonic function which is not absolutely continuous and let $\{c_n(F)\}$ be its Fourier-Stieltjes coefficients

$$c_n(F) = \int_{-\pi}^{\pi} e^{inx} dF(x).$$

It has been shown by Littlewood [1] that there exists an $F(x)$ and a positive constant α such that

$$c_n(F) = O(|n|^{-\alpha}).$$

In a recent paper Wiener and Wintner [2] have shown that α may be any number up to $\frac{1}{2}$, that is, for every $\epsilon > 0$ there exists an $F(x)$, depending on ϵ, such that

$$c_n(F) = O(|n|^{-\frac{1}{2}+\epsilon}).$$

Let $r(n)$, defined for positive integer values of the argument, be a positive function which becomes infinite, however slowly, as n becomes infinite. In this paper, using the method of Wiener and Wintner, their result is sharpened by showing that for given $r(n)$ there exists an $F(x)$ such that

$$(1) \qquad c_n(F) = O(r(|n|)\,|n|^{-\frac{1}{2}}), \qquad n \to \pm\infty.$$

It is first shown that for given $r(n)$ there exists an $F^*(x)$, continuous and monotonic, but not absolutely continuous, such that all but a sufficiently small proportion of its Fourier-Stieltjes coefficients vanish. This will imply that their average value is small,

$$\sum_{n=-N}^{N} |c_n(F^*)| = O(r(N)).$$

With the aid of $F^*(x)$ it is then possible to construct an $F(x)$, continuous and monotonic, but not absolutely continuous, whose coefficients satisfy (1). This function satisfies the additional condition, as have earlier examples, that

* Received January 5, 1939.

[1] J. E. Littlewood, "On Fourier coefficients of functions of bounded variation," *Quarterly Journal of Mathematics*, vol. 7 (1936), pp. 219-226.

[2] Norbert Wiener and Aurel Wintner, "Fourier-Stieltjes transforms and singular infinite convolutions," *American Journal of Mathematics*, vol. 40 (1938), pp. 513-522.

its derivative vanishes almost everywhere. There will be no loss of generality in supposing that $r(n)$ is an increasing function of n.

LEMMA. *Let $r(n)$ be a positive increasing function of n which tends to infinity as n tends to infinity*

$$0 < r(1) < r(2) < r(3) < \cdots$$
$$r(n) \to \infty \text{ as } n \to \infty.$$

Then there is a monotonic singular (i. e., continuous and with derivative equal to zero almost everywhere, but not constant) function, $P(x)$, such that

$$\sum_{n=-2k}^{2k} | c_n(P) | = O(r(k)). \tag{2}$$

Proof. Let $r(n)$ satisfy the conditions of the Lemma and let $m_1, m_2, m_3, \cdots$ be a sequence of positive integers for which

$$\left.\begin{aligned} r(m_k) &> 2^k \\ m_{k+1} &> 8^{k+1} m_k \end{aligned}\right\} k = 1, 2, 3, \cdots . \tag{3}$$
$$m_1 > 8$$

It is to be shown that the function [3]

$$P(x) = \lim_{n\to\infty} \int_0^x \prod_{\nu=1}^{n} (1 + \cos m_\nu t)\, dt \tag{4}$$

satisfies the requirements of the Lemma. Letting $g_n(t)$ denote the non-negative integrand we see that $g_n(x)$ is a cosine polynomial of degree γ_n where

$$\gamma_n = m_1 + m_2 + \cdots + m_n.$$

In virtue of (3)

$$m_n \leqq \gamma_n < 2m_n. \tag{5}$$

If

$$g_n(x) = 1 + \sum_{\nu=1}^{\gamma n} a_\nu \cos \nu x$$

the relation $g_{n+1}(x) - g_n(x) = \cos m_{n+1} x g_n(x)$ shows that

$$g_{n+1}(x) - g_n(x) = \cos m_{n+1} x + \tfrac{1}{2} \sum_{\nu=1}^{\gamma n} a_\nu \{\cos (m_{n+1} - \nu) x + \cos (m_{n+1} + \nu) x\} \tag{6}$$

so $g_{n+1}(x) - g_n(x)$ is a sum of cosine terms each of which is of degree at

---

[3] Compare F. Riesz, "Über die Fourierkoeffizienten einer stetigen Function von beschränkter Schwankung," *Mathematische Zeitschrift*, vol. 2 (1918), pp. 312-315.

least $m_{n+1}-\gamma_n$ and by (3) and (5) this is greater than $8m_n-\gamma_n>\gamma_n$. It follows that $g_{n+1}(x)$ is obtained from $g_n(x)$ by the addition of terms of higher order than any in $g_n(x)$, and these cosine polynomials are therefore partial sums of a series

$$(7)\qquad 1+\sum_{\nu=1}^{\infty} a_\nu \cos \nu x.$$

Let s_n denote the sum of the absolute magnitude of all coefficients of $g_n(x)$ (actually $a_\nu \geqq 0$),

$$s_n = 1+\sum_{\nu=1}^{\gamma_n} |a_\nu|.$$

Then using (6) one shows by induction that

$$s_n = 2^n.$$

If $P_n(x)$ denotes the integral of $g_n(t)$ from 0 to x then $P_n(x)$ is a partial sum of the series

$$(8)\qquad x+\sum_{\nu=1}^{\infty} \frac{a_\nu}{\nu}\sin \nu x$$

which is obtained by formally integrating (7), and $P_n(x)$ is non-decreasing since $g_n(t)$ is non-negative. To show that this series converges absolutely we observe from (6) that in the sum $\sum_{\nu=\gamma_n+1}^{\gamma_{n+1}} \left|\frac{a_\nu}{\nu}\right|$ all the coefficients are zero for which ν is less than $m_{n+1}-\gamma_n$. Thus

$$(9)\qquad \sum_{\gamma_n+1}^{\gamma_{n+1}} \left|\frac{a_\nu}{\nu}\right| < \frac{1}{m_{n+1}-\gamma_n}\sum \left| a_\nu \right| < \frac{2^{n+1}}{m_{n+1}} < \frac{2^{n+1}}{8^{n+1}}$$

since by (3) and (5) $\gamma_n < 2m_n < \frac{1}{4}m_{n+1}$. Thus by grouping terms in the series (8) for which ν lies between γ_n+1 and γ_{n+1}, $n=1,2,3,\cdots$, we find that (8) converges absolutely, and therefore uniformly.

Thus the series (8) is the uniform limit as n becomes infinite of $P_n(x)$ and defines a continuous non-decreasing function. The sum of this series must be $P(x)$, for by definition, (4), $P(x)=\lim_{n\to\infty} P_n(x)$.

It is easily verified that

$$\int_{-\pi}^{\pi} e^{inx}dP(x) = \begin{cases} \pi a_{|n|};\ n=\pm 1, \pm 2, \cdots \\ 2\pi;\ n=0. \end{cases}$$

If k is any integer greater than γ_1 we can write $\gamma_n < k \leqq \gamma_{n+1}$ and then $2k$ will be less than γ_{n+2}. From (3) and the fact that $r(n)$ is increasing we have

$$1+\sum_{\nu=1}^{2k} |a_\nu| < s_{n+2} = 2^{n+2} < 4r(m_n) < 4r(k).$$

Then for all large k

$$\sum_{\nu=-2k}^{2k} | c_\nu(P) | < 8\pi r(k).$$

The Fourier-Stieltjes coefficients of $P(x)$ are πa_n and since a_n does not tend to zero as n tends to infinity ($a_{m_\nu} = 1$; $\nu = 1, 2, 3, \cdots$) it follows from the Riemann-Lebesgue Theorem that $P(x)$ is not absolutely continuous. But slightly more than this is true: we shall show that the derivative of $P(x)$ is equal to zero almost everywhere. It is first shown that for every x

$$\frac{P(x+h) - P(x-h)}{2h} - \prod_{\nu=1}^{n} (1 + \cos m_\nu x) \tag{10}$$

tends to zero as n tends to infinity and h tends to zero through a sequence of values depending on n. Remembering that the product is a partial sum of series (7) and $P(x)$ is the sum of the formally integrated series (8) we see that (10) is equal to

$$\sum_{\nu=1}^{\gamma_n} a_\nu \cos \nu x \left(\frac{\sin \nu h}{\nu h} - 1\right) + \sum_{\nu=\gamma_n+1}^{\infty} \frac{a_\nu}{\nu h} \cos \nu x \sin \nu h. \tag{11}$$

Let

$$h = \frac{1}{2^n \gamma_n}.$$

On grouping terms in the second sum of (11) which lie between $\gamma_j + 1$ and γ_{j+1}; $j = n, n+1, n+2, \cdots$; we see from (9) that this sum is less than

$$\frac{1}{h}\left(\frac{2^{n+1}}{m_{n+1}} + \frac{2^{n+2}}{m_{n+2}} + \frac{2^{n+3}}{m_{n+3}} + \cdots \right).$$

Since, by (3), $m_{n+\nu} > 8^{n+\nu} m_n$ this sum is less than

$$\frac{1}{3hm_n4^n} < \frac{1}{h\gamma_n 4^n} = \frac{2^n}{4^n}.$$

From the elementary inequality $\left| \dfrac{\sin x}{x} - 1 \right| < x^2$, $0 < x < 1$, it follows that the first sum of (11) is less than

$$\sum_{\nu=1}^{\gamma_n} | a_\nu | \nu^2 h^2 < (h\gamma_n)^2 \sum | a_\nu | < \frac{2^n}{4^n}.$$

Thus (10) approaches zero for every x, but the first term of (10) approaches $P'(x)$ which exists and is finite except over a set of measure zero since $P(x)$ is non-decreasing. Then as n tends to infinity the product

$$\prod_{\nu=1}^{n} (1 + \cos m_\nu x)$$

tends to $P'(x)$, which is finite almost everywhere. If this non-negative product tends to a finite limit different from zero in a set E then $\cos m_n x$ tends to zero in the set E as $m_n \to \infty$ and by the Cantor-Lebesgue Theorem the measure of E is zero. Thus $P'(x)$ is equal to zero almost everywhere and $P(x)$ satisfies all the requirements of the Lemma.

THEOREM. *Let $r(n)$ satisfy the conditions of the Lemma. There is a monotonic singular function $F(x)$ such that*

$$(1) \qquad \left| \int_{-\pi}^{\pi} e^{inx} dF(x) \right| = O(r(|n|)\,|n|^{-\frac{1}{2}}), \quad n \to \pm\infty.$$

Proof. Following Wiener and Wintner we write

$$(12) \qquad \begin{aligned} y(x) &= \tfrac{1}{4}(3x + \frac{1}{\pi} x^2 \operatorname{sgn} x) \\ F(y) &= P(x), \ -\pi \leqq x \leqq \pi \end{aligned}$$

where $P(x)$ is the function constructed in the proof of the Lemma. The transformation (12) maps the interval $(-\pi, \pi)$ onto itself in a $1:1$ manner and since $\frac{dy}{dx}$ has a positive upper and lower bound in this interval the property of being monotonic and singular which is possessed by $P(x)$ must also be possessed by $F(y)$. The coefficients of $F(y)$ are related to $P(x)$ by

$$(13) \qquad c_k(F) = \int_{-\pi}^{\pi} e^{iky} dF(y) = \int_{-\pi}^{\pi} e^{iky(x)} dP(x).$$

Let $e^{iky(x)}$ be developed in the Fourier Series

$$(14) \qquad e^{iky(x)} = \sum_{n=-\infty}^{\infty} \lambda_{k,n} e^{inx}, \qquad -\pi \leqq x \leqq \pi.$$

By means of the second mean value theorem and a lemma of Van der Corput, Wiener and Wintner[2] show that

$$(15) \qquad |\lambda_{k,n}| < \begin{cases} \dfrac{A}{\sqrt{|k|}} & \text{for all } k \text{ and } n \\ \dfrac{A}{n^2} & \text{for } |n| > 2|k| \end{cases}$$

where A is independent of k and n. Thus series (14) converges uniformly for each k and we may substitute in (13) and integrate termwise, obtaining

$$c_k(F) = \sum_{n=-\infty}^{\infty} \lambda_{k,n} c_n(P).$$

Then

$$| c_k(F) | \leqq \sum_{|n| \leqq 2|k|} | \lambda_{k,n} c_n(P) | + \sum_{|n| > 2|k|} | \lambda_{k,n} c_n(P) |.$$

Since $P(x)$ is of bounded variation the coefficients $c_n(P)$ are uniformly bounded, and using inequalities (2) and (15) the first sum is less than

$$A \mid k \mid^{-\frac{1}{2}} \sum \mid c_n(P) \mid = O(\mid k \mid^{-\frac{1}{2}} r(\mid k \mid)).$$

The second sum is

$$O(1) \sum_{n > 2|k|} \frac{1}{n^2} = O(\mid k \mid^{-1}).$$

Thus

$$c_k(F) = O(\mid k \mid^{-\frac{1}{2}} r(\mid k \mid))$$

which completes the proof of the Theorem.

The Fourier-Stieltjes transform of the function $F(y)$ tends to zero, although perhaps not as rapidly as the sequence of coefficients. It can be shown [4] that

$$\int_{-\pi}^{\pi} e^{\pm iux} dF(x) = O(u^{-\frac{1}{2}} r(u) \log u), \qquad u \to \infty. \tag{16}$$

To this end we note that since $P(x)$ is the sum of series (8) if $\phi(u)$ is its transform

$$\phi(u) = \int_{-\pi}^{\pi} e^{iux} dP(x) = 2 \sin \pi u \left(\frac{1}{u} + \sum_{\nu=1}^{\infty} \frac{(-1)^{\nu} a_{\nu} u}{u^2 - \nu^2} \right)$$

so if u lies between λ and $\lambda + 1$ where λ is a large positive integer

$$| \phi(u) | < A_1 \left(\frac{1}{\lambda} + \sum_{\nu=1}^{2\lambda} \frac{a_\nu}{| \lambda - \nu | + 1} + \sum_{2\lambda+1}^{\infty} \frac{a_\nu}{\nu} \right).$$

From inequalities (9) and (3) we find that

$$\sum_{\lambda=1}^{2N} \sum_{\nu=2\lambda+1}^{\infty} \frac{a_\nu}{\nu} = O(r(N)).$$

Then, remembering that $\phi(u)$ is bounded and even, it follows that

$$\sum_{\lambda=-2N-1}^{2N} \max_{\lambda \leqq u \leqq \lambda+1} | \phi(u) | = O(r(N) \log N). \tag{17}$$

---

[4] To extend the definition of $r(u)$ to non-integer values of u we suppose that $r(u)$ is a linear function in the intervals $n \leqq u \leqq n + 1$; $n = 1, 2, 3, \cdots$.

Since $F(y)$ is defined by (12)

$$\int_{-\pi}^{\pi} e^{iuy} dF(y) = \int_{-\pi}^{\pi} e^{iuy(x)} dP(x) \tag{13a}$$

and now we develop $e^{iuy(x)}$ in a Fourier Series over an interval slightly greater than 2π in order that it be continuous in the closed interval. Let α be the smallest number which is equal to or greater than 1 and such that $uy(\alpha\pi)$ is an integral multiple of π, and let

$$e^{iuy(x)} = \sum_{n=-\infty}^{\infty} \lambda_{u,n} e^{inx/a}, \qquad -\alpha\pi \leqq x \leqq \alpha\pi. \tag{14a}$$

To prove (16) it is sufficient to suppose that u is positive since the transform of $F(y)$ is an even function of u. If u is large and positive the inequality (15) remains valid with k replaced by u, so substituting (14a) in (13a) and integrating termwise we obtain with the aid of (15) and (17) the estimate (16).

It is possible to further specialize functions of bounded variation. Wiener and Wintner show that their estimate on the magnitude of the coefficients is possible both for monotonic singular functions which are strictly increasing and for monotonic singular functions which are of the Cantor[5] type (i. e., constant over a set of intervals of length 2π). The function constructed in this paper is strictly increasing or is easily made so by a slight change in construction.

PURDUE UNIVERSITY.

---

[5] Norbert Wiener and Aurel Wintner, "On singular distributions," *Journal of Mathematics and Physics*, vol. 17 (1938), pp. 233-246.

A CLASS OF MONOTONE FUNCTIONS.*

By G. Baley Price.

1. Introduction. A function $x(t)$ may be said to be monotone if and only if $x(t)$ is between $x(t_1)$ and $x(t_2)$ whenever t is between t_1 and t_2. A wide variety of monotone functions in this sense results from the introduction of different types of betweenness. In a linear space with elements u, v, w it is natural to say that w is between u and v if and only if $w = \theta u + (1 - \theta) v$, where θ is a real number such that $0 \leqq \theta \leqq 1$. Let $x(t)$ be a function defined on a convex set in a linear space with values in a similar space. If $x(t)$ is monotone in the sense stated with betweenness interpreted as explained, we shall say that it is linear-monotone.

This note treats real-valued linear-monotone functions $x(s, t)$ defined on a closed, convex set C of the euclidean plane. Thus $x(s, t)$ is a real-valued function of two independent variables which is monotone in the ordinary sense along every straight line in its region C of definition. The results obtained include a complete determination of the nature and distribution of the points of discontinuity of $x(s, t)$, and a complete description of $x(s, t)$ in the large. We assume that $x(s, t)$ has no removable discontinuities.

2. Two lemmas. Some preliminary results are needed for an analysis of the discontinuities of $x(s, t)$.

(2.1) Lemma. *Let $u(\theta)$ be a real-valued function defined on the circumference C^* of the unit circle with center P, and let $u(\theta)$ be monotone on every open semi-circle of C^*. Then $u(\theta)$ has one of the following two forms:* (a) *$u(\theta)$ is constant on C^**; (b) *there exists a line L through P which divides C^* into two open semi-circles C^*_1, C^*_2 at points P_1, P_2; $u(\theta)$ is constant on C^*_1 and equal to u_M, its maximum value; $u(\theta)$ is constant on C^*_2 and equal to u_m; at P_1, P_2 the values of $u(\theta)$ are numbers on the interval $u_m \leqq u \leqq u_M$.*

If $u(\theta)$ is not a constant, then either (i) $u(\theta)$ has a maximum or (ii) $u(\theta)$ has a point of discontinuity; in both cases we shall show that $u(\theta)$ has the form stated in (b).

(i) There is an open semi-circle C^*_1, with end-points P_1 and P_2, on which $u(\theta)$ takes on its maximum value u_M; the assumption that it takes on its maximum value on a set less than an open semi-circle or more than a semi-

* Received October 4, 1938; revised May 5, 1939.

circle but less than the entire circle contradicts the fact that $u(\theta)$ is monotone on every open semi-circle. Let θ_1 be any value of θ on the complementary open semi-circle C^*_2. It follows that $u(\theta) = u(\theta_1)$ for θ on C^*_2, and that $u(\theta)$ has the form stated in (b).

(ii) Let $P_1 : \theta_1$ be a point of discontinuity of $u(\theta)$. Because of the monotone character of $u(\theta)$, the limits $u(\theta_1 - 0)$, $u(\theta_1 + 0)$ exist; we may assume that the second is greater than the first without loss of generality. We observe first that $u(\theta) \geqq u(\theta_1 + 0)$ on $C^*_1 : \theta_1 < \theta < \theta_1 + \pi$, next that $u(\theta) = u(\theta_1 - 0)$ on C^*_2: $\theta_1 + \pi < \theta < \theta_1 + 2\pi$, and finally that $u(\theta) \leqq u(\theta_1 + 0)$ on C^*_1. The two inequalities show that $u(\theta) = u(\theta_1 + 0)$ on C^*_1. The complete proof follows.

There is an extension of this lemma. Let $u(s)$ be a real-valued function, defined for s on the surface of the unit sphere S: $t_1^2 + t_2^2 + t_3^2 = 1$, which is monotone on every open semi-circle of a great circle. Then $u(s)$ has one of the following two forms: (a) $u(s)$ is constant on S; (b) there exist two open hemi-spheres S_1, S_2 separated by a great circle C^*; $u(s)$ is constant and equal to u_M, its maximum value, on S_1; $u(s)$ is constant on S_2 and equal to u_m; finally, on C^* $u_m \leqq u(s) \leqq u_M$, and $u(s)$ has the form stated in Lemma 2.1. On every great circle of S, $u(s)$ is a function of the kind described in Lemma 2.1. If there exist two great circles of S on which $u(s)$ is constant, $u(s)$ is constant over all of S; $u(s)$ then has the form stated in (a). In all other cases there exist at least two great circles on which $u(s)$ is discontinuous. Then the great circle C^* through the four points of discontinuity divides S into open hemi-spheres S_1 and S_2, and $u(s)$ has the form stated in (b).

Consider again the real-valued, linear-monotone function $x(s, t)$ defined on the closed convex set C. Let P be any point of C, and let $x^*(r, \theta)$ be the function $x(s, t)$ expressed in terms of polar coördinates with P as pole. Let $u(\theta)$ be defined by

$$u(\theta) = \lim_{r \to 0} x^*(r, \theta). \tag{2.2}$$

It is convenient to think of $u(\theta)$ as defined on the circumference of the unit circle C^* with P as center.

(2.3) LEMMA. *The function $u(\theta)$ in (2.2) is monotone on every open semi-circle of C^* or part of an open semi-circle on which it is defined. If P is an interior point of C, $u(\theta)$ has the form* (a) *or* (b) *of Lemma* 2.1.

The limit of a sequence of monotone functions is a monotone function, and $x(s, t)$ is an ordinary monotone function along every straight line in C. The lemma follows from these facts and (2.2).

3. Types of discontinuities. Let $u(\theta)$ be the function defined in (2.2),

and let $\theta_1 \leqq \theta \leqq \theta_2$ be an arc of C^* on which $u(\theta)$ is constant. Then $\lim x(s, t)$ exists as a two dimensional limit as (s, t) approaches P in the closed sector subtended by the arc $\theta_1 \leqq \theta \leqq \theta_2$. Thus if $u(\theta)$ is constant at an interior point P of C, $x(s, t)$ is continuous there; otherwise, $u(\theta)$ takes on at most four values, and we shall say that $x(s, t)$ has a simple discontinuity at P. With each point of discontinuity P on the interior of C there is associated the line L of Lemma 2. 1; along any other line through P $x(s, t)$ has a jump there equal to $u_M - u_m$. If P is a boundary point of C, it is either a point of continuity or $u(\theta)$ is a monotone function which is not constant. A discontinuity on the boundary of C will be called a helical discontinuity for obvious reasons.

Let P: (s_0, t_0) be an interior point of C. If $x(s, t_0)$ is continuous in s for $s = s_0$, and if $x(s_0, t)$ is continuous in t for $t = t_0$, then $x(s, t)$ is continuous at P. For if $x(s, t)$ were discontinuous at P, there would be a line L across which it would have a jump, and at least one of the functions $x(s, t_0)$, $x(s_0, t)$ would be discontinuous.

4. Distribution of points of discontinuity. Let $x(s, t)$ have a simple discontinuity at P in C; let L be the line through P across which the jump occurs; and let S be the open segment of L in C. Denote the two convex sets into which L divides C by C_1 and C_2. Let the limits at P along rays in C_1, C_2 be $u_m(P), u_M(P)$ respectively. Denote by P^* any point of S distinct from P. The limit of $x(s, t)$ at P^* along any ray in C_1 is equal to or less than $u_m(P)$, and along any ray in C_2 it is equal to or greater than $u_M(P)$. Thus P^* is a point of discontinuity of $x(s, t)$, and L is the associated line across which the jump occurs. Then there are two limits $u_m(P^*)$, $u_M(P^*)$ at P^* along rays in C_1, C_2 respectively with $u_m(P^*) \leqq u_m(P)$, $u_M(P^*) \geqq u_M(P)$. By starting with P^* and considering the limits at P, we obtain $u_m(P) \leqq u_m(P^*)$, $u_M(P) \geqq u_M(P^*)$. We thus have the following theorem.

(4. 1) THEOREM.[1] *$x(s, t)$ has no isolated points of discontinuity on the interior of C. If $x(s, t)$ has a simple discontinuity at P, and if L is the line through P across which the jump occurs, then every point of the open segment S of L in C is a simple discontinuity with the same jump across L. The end-points of S are helical discontinuities.*

Associated with each segment S of discontinuities there are two limits $u_m(S)$ and $u_M(S)$. If we agree that S_1 precedes S_2 if and only if $u_m(S_1) < u_M(S_1) \leqq u_m(S_2) < u_M(S_2)$, the segments of discontinuities form a simply

[1] The author is indebted to Professor Saunders MacLane for the proof of this theorem.

ordered class. This ordering has an important geometric characteristic. Let S, S_1, S_2 be any three segments with ends on the boundary of C. If S_1 lies in one of the regions into which S divides C and S_2 lies in the other, we shall say that S is geometrically between S_1 and S_2. If S, S_1, S_2 are segments of discontinuities of $x(s,t)$, and if S is between S_1, S_2 in the ordering just defined, then S is geometrically between S_1 and S_2. For the assumption that S is not geometrically between S_1 and S_2 leads to the conclusion that $x(s,t)$ is not monotone along every straight line in C, which contradicts the hypotheses. We thus have the following theorem.

(4.2) THEOREM. *The segments of simple discontinuities of $x(s,t)$ form a simply ordered class in which betweenness is geometric betweenness.*

5. Description of the function in the large. From Theorem 4.2 we have the following theorem.

(5.1) THEOREM. *The set of segments of simple discontinuities of $x(s,t)$ is denumerable.*

The following lemma is needed for the proof of the next theorem; its proof is omitted.

(5.2) LEMMA. *Let C be a closed convex set, and let C_1, C_2 be two disjoint convex sets whose point-set sum is C. Then the common boundary of C_1, C_2 is a closed segment of a straight line.*

(5.3) THEOREM. *The closed convex set C can be covered by a simply ordered set of segments of lines in which betweenness is geometric betweenness. The segments of simple discontinuities of $x(s,t)$ form a subset of the segments of this covering, and $x(s,t)$ is constant at least on the interior of each of the remaining segments.*

It has been shown already that the segments of simple discontinuities form a simply ordered set. Let $P\colon (s_0, t_0)$ be an interior point of C which lies between two consecutive segments S_1, S_2 of discontinuities. Since $x(s,t)$ is continuous on the interior of C between S_1 and S_2, the points at which $x(s,t) = x(s_0,t_0)$ separate S_1 and S_2. The sets C_1 and C_2 in which $x(s,t) < x(s_0,t_0)$ and $x(s,t) \geqq x(s_0,t_0)$ respectively are convex and their sum is C; by Lemma 5.2 their common boundary is a segment of a straight line. Likewise, the boundary of the regions in which $x(s,t) \leqq x(s_0,t_0)$ and $x(s,t) > x(s_0,t_0)$ is a segment of a line. Hence, the region of C in which $x(s,t) = x(s_0,t_0)$ is bounded by two arcs of the boundary of C and by two segments of straight lines; in particular, it may be a segment of a single straight line. These are the essential facts in the proof of the theorem, the details of which are left to the reader.

Theorem 5.3 shows that the surface corresponding to the function $x(s,t)$ is part of a ruled surface except for discontinuities along a denumerable set of segments in C.

(5.4) THEOREM. *A necessary and sufficient condition that a point P on the boundary of C be a helical discontinuity is that it be either an end-point of a segment of simple discontinuities or the point of intersection of two segments on the interiors of which $x(s,t)$ is constant with unequal values.*

The condition stated is obviously sufficient. To prove that it is necessary, assume that P is a helical discontinuity but does not have the character stated. In the general case then there is a first segment PQ and a last segment PR through P such that $x(s,t) = c$, a constant, on the open segments PQ, PR and on the open segments in C of all lines through P in the angle QPR. Then since PQ and PR are not segments of simple discontinuities, there exists a segment S_1 preceding PQ and a segment S_2 following PR such that $c - \epsilon < x(s,t) < c + \epsilon$ in the part of C bounded by S_1 and S_2. Since we are assuming the theorem false, S_1 and S_2 do not intersect at P. Then $x(s,t)$ is continuous at P, and this contradiction establishes the theorem.

(5.5) THEOREM. *The helical discontinuities are at most denumerable in number.*

By Theorems 5.1 and 5.4 there are at most a denumerable number of helical discontinuities corresponding to segments of simple discontinuities of $x(s,t)$. Associated with any other helical discontinuity there is a certain increase in the function; it follows that the set of all such discontinuities is denumerable. The proof is complete.

Let S_1, S_2 be two consecutive segments of simple discontinuities of $x(s,t)$ which cut off arcs A, B of the boundary of C. The direction of increase of $x(s,t)$ is the same for every helical discontinuity on A; also it is the same for every helical discontinuity on B but opposite to that for those on A.

(5.6) THEOREM. *If $x(s,t)$ has a maximum value on a closed convex set C' in C, there is an extreme point of C' at which $x(s,t)$ takes on this maximum value. A similar statement holds for a minimum value.*

Assume that $x(s,t)$ has the maximum value x^* on C', and assume that $x(s,t) < x^*$ at every extreme point of C'. Then since C' is the convex extension of its extreme points (see [4, Theorem 9.1, p. 63]), and since $x(s,t)$ is linear-monotone on C', it follows that $x(s,t) < x^*$ at all points of C'. This contradiction establishes the theorem. Examples show that $x(s,t)$ may have neither a maximum nor a minimum.

6. Conclusion. A real-valued function of two variables $u(s,t)$ is monotonic in a region R in the sense of Lebesgue [2, p. 385] and McShane [3, p. 717] if in every open subset R^* of R it satisfies the inequality $m^*_b \leqq u(s,t) \leqq M^*_b$, where m^*_b, M^*_b denote the greatest lower and least upper bounds respectively of $x(s,t)$ on the boundary of R^*. The linear-monotone function $x(s,t)$ is monotonic on C in the sense of Lebesgue and McShane.

Numerous examples of linear-monotone functions $x(s,t)$ can be given. Furthermore, examples can be constructed to show that a monotone function in the sense of Arzelà [1, p. 343] may not be linear-monotone, and others to show that a linear-monotone function may not be monotone in the sense of Arzelà.

By prescribing the behavior of $x(s,t)$ along every line in C, we have prescribed its behavior in the large. The interval of definition of an ordinary monotone function can be enlarged, but the region of definition of $x(s,t)$ may have a natural boundary since a helical discontinuity cannot occur at an interior point. The only function $x(s,t)$ defined over the entire plane is essentially a function of a single variable. In several respects, therefore, $x(s,t)$ is similar to an analytic function.

The functions in the linear extension of the set of linear-monotone functions $x(s,t)$ on C are characterized by discontinuities of simple types. These can be determined from the results established above.

The extensions of these results to real-valued linear-monotone functions $x(t_1, t_2, t_3)$ of three variables will be left to the reader. The methods and results are similar to those above. The fundamental result for the analysis of points of discontinuity is contained in the extension of Lemma 2.1 which was given in section 2.

THE UNIVERSITY OF KANSAS,
LAWRENCE, KANSAS.

REFERENCES.

1. E. W. Hobson, *The Theory of Functions of a Real Variable*, vol. 1, 3rd edition (1927), Cambridge University Press.

2. H. Lebesgue, "Sur le Problème de Dirichlet," *Rendiconti del Circolo Matematico di Palermo*, vol. 24 (1907), pp. 371-402.

3. E. J. McShane, "Parametrizations of Saddle Surfaces with Applications to the Problem of Plateau," *Transactions of the American Mathematical Society*, vol. 35 (1933), pp. 716-733.

4. G. B. Price, "On the Extreme Points of Convex Sets," *Duke Mathematical Journal*, vol. 3 (1937), pp. 56-67.

ASYMPTOTIC PROBABILITY DISTRIBUTIONS AND HARMONIC CURVES.*

By E. K. Haviland.

The mechanical production of harmonic curves, at least by such devices as the double pendulum and the simpler geared machines, is a familiar process, and a number of elaborate machines have been constructed capable of producing beautiful and intricate curves. The subject has been carefully treated by William F. Rigge, S. J., in his book, *Harmonic Curves*.[1] Since the fundamental operation in the production of these curves is one of the addition of given independent distributions on line segments or convex curves, it is of interest to examine these curves in the light of the mathematical theory of the addition of distributions on convex curves, a subject which has been treated by a number of authors [2] in recent years. While the problem is originally one concerning asymptotic distributions, it reduces to a spatial convolution problem in virtue of the Kronecker-Weyl Theorem. It is the purpose of the present note, by the investigation of the addition of given independent probability distributions on two ellipses, on the one hand to provide a mathematical theory for some very interesting experimental results and on the other hand to illustrate the general mathematical theory of the addition of independent distributions on convex curves by the treatment of a case in which the situations arising can readily be visualized and in which explicit formulae and criteria can be obtained.

If two closed convex curves, S_1 and S_2, in the z-plane, where $z = x + iy$, be represented parametrically in terms of ϑ_1, ϑ_2 respectively, then the vector addition of the curves corresponds to a mapping of the $\vartheta_1 \times \vartheta_2$-torus on a

* Received January 5, 1939.

[1] William F. Rigge, S. J., *Harmonic Curves*, The Creighton University, Omaha, Nebraska, 1926.

[2] Cf., e. g., H. Bohr, "Om Addition af uendelig mange konvekse Kurver," *Danske Videnskabernes Selskab*, Forhandlinger, 1913, pp. 325-366; H. Bohr and B. Jessen, "Om Sandsynlighedsfordelinger ved Addition af konvekse Kurver," *Danske Videnskabernes Selskab*, Skrifter (8), vol. 12, no. 3 (1929); E. K. Haviland, "On the addition of convex curves in Bohr's theory of Dirichlet series," *American Journal of Mathematics*, vol. 55 (1933), pp. 332-334; R. B. Kershner, "On the addition of convex curves," *American Journal of Mathematics*, I, vol. 58 (1936), pp. 737-746; II, vol. 59 (1937), pp. 423-426; E. R. van Kampen and A. Wintner, "Convolutions of distributions on convex curves and the Riemann zeta function," *American Journal of Mathematics*, vol. 59 (1937), pp. 175-204.

region of the z-plane which will be denoted by $S_1(+)S_2$. It is well known that $S_1(+)S_2$ is either a region bounded by a single convex curve, C, or else an annular region bounded by two convex curves, C and D, where D lies within C. To the points on the torus at which the Jacobian of the mapping function vanishes, there correspond in the z-plane two curves, Γ_1 and Γ_2, of which the former is identical with C, while the latter contains D, if D exists, and may coincide with D. Γ_1 and Γ_2 divide $S_1(+)S_2$ into connected subregions characterized by different densities of distribution of the mapping function, in a manner which will be indicated more fully below. In this connection, the curves shown by Rigge in figure 962, opposite page 176, of his *Harmonic Curves* are of particular interest. With the exception of those in the fourth row and those in the fourth and tenth columns, they arise from the vector addition of ellipses, being, in fact, the maps in the z-plane of certain straight lines on the torus. All represent cases where the inner curve, D, does not exist, but in many of them the location of the curve Γ_2 can readily be traced, and the different densities of the distribution in the connected subregions of $S_1(+)S_2$ is clearly indicated.

Turning now to the more detailed treatment of our problem, we let the function

$$x + iy = z = z_j(\vartheta_j) = \xi_j(\vartheta_j) + i\eta_j(\vartheta_j), \qquad (j = 1, 2), \tag{1}$$

where ϑ_j is the angular coördinate on the oriented circle Θ_j, be an admissible parametric representation, T_j, of the convex curve S_j in the z-plane. It may be assumed that the orientation of Θ_j, when transplanted by means of T_j onto S_j, determines on S_j that orientation which is positive in the z-plane. Let T denote the transformation

$$x + iy = z(\vartheta_1, \vartheta_2) = \sum_{j=1}^{2} z_j(\vartheta_j) = \sum_{j=1}^{2} \xi_j(\vartheta_j) + i \sum_{j=1}^{2} \eta_j(\vartheta_j) \tag{2}$$

of the torus $\Theta_1 \times \Theta_2$ into a subset of the z-plane, viz., the vector sum $S_1(+)S_2$, of the convex curves S_1, S_2. Then if

$$J = J(\vartheta_1, \vartheta_2) = \frac{\partial(x, y)}{\partial(\vartheta_1, \vartheta_2)} = \begin{vmatrix} x'_1(\vartheta_1) & x'_2(\vartheta_2) \\ y'_1(\vartheta_1) & y'_2(\vartheta_2) \end{vmatrix} \tag{3}$$

denote the Jacobian of the transformation T, van Kampen and Wintner have shown [3] that the set of those points on the torus $\Theta_1 \times \Theta_2$ at which the Jacobian (3) vanishes consists of two disjoint rectifiable Jordan curves, Ω^+ and Ω^-, corresponding respectively to those points at which the oriented normal to S_1 at the T_1-image of the point ϑ_1 of Θ_1 is either parallel or antiparallel to the

[3] *Loc. cit.*, pp. 179-182.

oriented normal to S_2 at the T_2-image of the point ϑ_2 of Θ_2. Furthermore, if Γ_1, Γ_2 denote the T-images of the curves Ω^+, Ω^-, then, as indicated above, Γ_1 is identical with outer boundary, C, of $S_1(+)S_2$, while the inner boundary, D, if it exists, is contained in, and may or may not coincide with, the curve Γ_2, coincidence occurring if and only if D is free of corners. If we put

$$R = [S_1(+)S_2] - [\Gamma_1 + \Gamma_2], \tag{4}$$

and if $z = z_0$ is any fixed point of the (open) set R and $\epsilon = \epsilon(z_0) > 0$ is so small that the circle $U: |z - z_0| \leq \epsilon$ is contained in R, then there exist on the torus $\Theta_1 \times \Theta_2$ a finite number, $m = m(U)$, of mutually disjoint closed sets, $\Lambda_1, \cdots, \Lambda_k, \cdots, \Lambda_m$, such that every Λ_k is transformed by T into U in a topological way, while no point of $\Theta_1 \times \Theta_2 - \sum_{k=1}^{m} \Lambda_k$ is transformed by T into a point of U. Finally, for $\psi_2(E)$, the distribution function of (2), van Kampen and Wintner have obtained the explicit representation

$$\psi_2(E) = \iint \delta_2(x, y)\, dx dy, \tag{5}$$

where

$$\delta_2(x, y) = \sum_{k=1}^{m} | 4\pi^2 J(\alpha_k(x, y), \beta_k(x, y)) |^{-1}, \tag{6}$$

if the point $z = x + iy$ is in R, and $\delta_2(x, y) = 0$ otherwise; $\vartheta_1 = \alpha_k(x, y)$, $\vartheta_2 = \beta_k(x, y)$, $(k = 1, 2, \cdots, m)$, denoting the topological transformation T^{-1} of U into Λ_k.

We now apply the foregoing general results to the vector addition of two ellipses,

$$S_j: \ (x_j = a_j \cos \vartheta_j + b_j \sin \vartheta_j;\ y_j = c_j \cos \vartheta_j + d_j \sin \vartheta_j), \quad (j = 1, 2), \tag{7}$$

where an explicit treatment is possible by means of elementary methods. In particular, the nature of the curve Γ_2 will be studied and conditions will be obtained for the existence of cusps on Γ_2 and also for the existence of the inner boundary, D, in terms of the constants $a_1, \cdots, d_2$; also values for m in the different parts of $S_1(+)S_2$ will be determined.

It is assumed that the ellipses are concentric, since this affects only the location of $S_1(+)S_2$ but not its form. Otherwise, in our general case, the ellipses are chosen quite arbitrarily. Here

$$\begin{aligned} J = {} & (a_1c_2 - a_2c_1) \sin \vartheta_1 \sin \vartheta_2 + (b_1d_2 - b_2d_1) \cos \vartheta_1 \cos \vartheta_2 \\ & + (b_2c_1 - a_1d_2) \sin \vartheta_1 \cos \vartheta_2 + (a_2d_1 - b_1c_2) \cos \vartheta_1 \sin \vartheta_2, \end{aligned} \tag{8}$$

from which it may be seen that $J = 0$ is an algebraic curve of fourth degree. In the simple case in which S_1 and S_2 are similar and coaxial ellipses, we have

$$S_j: \quad (x = a_j \cos \vartheta_j;\ y = ka_j \sin \vartheta_j), \qquad (j = 1, 2),$$

where we may without restriction suppose $a_1 \geqq a_2$. Here $J = 0$ reduces to $\sin \vartheta_1 \cos \vartheta_2 - \cos \vartheta_1 \sin \vartheta_2 = \sin(\vartheta_1 - \vartheta_2) = 0$ and the corresponding curves on the torus are

$$\Omega^+: \quad \vartheta_2 = \vartheta_1; \qquad \Omega^-: \quad \vartheta_2 = \vartheta_1 + \pi.$$

In the z-plane, $C = \Gamma_1$, where Γ_1 is given by $(x = (a_1 + a_2)\cos \vartheta_1;\ y = k(a_1 + a_2)\sin \vartheta_1)$ and $D = \Gamma_2$, where Γ_2 is expressed as $(x = (a_1 - a_2)\cos \vartheta_1;\ y = k(a_1 - a_2)\sin \vartheta_1)$. In this case, the inner curve D always exists, save that it reduces to a point if S_1 and S_2 are identical. R consists of a single region, in which, as will be shown in the treatment of the general case, $m = 2$.

We now return to the general case and note that by virtue of our assumptions regarding the mapping of Θ_j on S_j, we suppose

$$(9) \qquad \Delta_j = a_j d_j - b_j c_j > 0, \qquad (j = 1, 2).$$

Furthermore, under the assumptions[4] $\vartheta_1, \vartheta_2 \neq \tfrac{1}{2}(2k + 1)\pi$, we have on the torus on curve Ω^+:

$$(10) \qquad \vartheta_2 = \arctan \frac{(a_1 d_2 - b_2 c_1)\sin \vartheta_1 + (b_2 d_1 - b_1 d_2)\cos \vartheta_1}{(a_1 c_2 - a_2 c_1)\sin \vartheta_1 + (a_2 d_1 - b_1 c_2)\cos \vartheta_1}$$

$$= \arctan \frac{(a_1 d_2 - b_2 c_1)\tan \vartheta_1 + (b_2 d_1 - b_1 d_2)}{(a_1 c_2 - a_2 c_1)\tan \vartheta_1 + (a_2 d_1 - b_1 c_2)},$$

and on curve Ω^-:

$$(11) \qquad \vartheta_2 = \arctan \frac{(a_1 d_2 - b_2 c_1)\tan \vartheta_1 + (b_2 d_1 - b_1 d_2)}{(a_1 c_2 - a_2 c_1)\tan \vartheta_1 + (a_2 d_1 - b_1 c_2)} + \pi.$$

A straightforward calculation then shows that on both curves

$$(12) \qquad \frac{d\vartheta_2}{d\vartheta_1} = \frac{\Delta_1 \Delta_2}{[(a_1 c_2 - a_2 c_1)\sin \vartheta_1 + (a_2 d_1 - b_1 c_2)\cos \vartheta_1]^2 + [(a_1 d_2 - b_2 c_1)\sin \vartheta_1 + (b_2 d_1 - b_1 d_2)\cos \vartheta_1]^2},$$

which is always positive. Moreover, the denominator is always positive in consequence of (9). For the purpose of examining the curves Γ_1, Γ_2 in the z-plane, we introduce the abbreviations

$$(13) \qquad \alpha = a_1 c_2 - a_2 c_1; \quad \beta = b_2 d_1 - b_1 d_2; \quad \gamma = a_1 d_2 - b_2 c_1; \quad \delta = a_2 d_1 - b_1 c_2$$

and

$$(14) \qquad M = [\alpha \sin \vartheta_1 + \delta \cos \vartheta_1]^2 + [\gamma \sin \vartheta_1 + \beta \cos \vartheta_1]^2,$$

and write (10) in the form[5]

---

[4] It is not difficult to see the modifications which must be made if these assumptions are not fulfilled.

[5] With appropriate modifications if $\alpha = 0$ or $\gamma = 0$.

(10a) $$\vartheta_2 = \arctan[(\gamma/\alpha)(\tan\vartheta_1 + \beta/\gamma)/(\tan\vartheta_1 + \delta/\alpha)]$$

and for definiteness assume

(15) $$\alpha, \beta, \gamma, \delta > 0,$$

assumptions which introduce no essential restrictions, as they merely determine the course of the curves Ω^+, Ω^- on the torus. From (9), it follows that

$$d\vartheta_2/d\vartheta_1 > 0, \quad \delta/\alpha > \beta/\gamma, \quad \gamma/\alpha > \beta/\delta.$$

Then on Ω^+, when $\vartheta_1 = 0$, $\vartheta_2 = \arctan(\beta/\delta)$ (in the first quadrant, we may suppose). As ϑ_1 increases from 0 to $\frac{1}{2}\pi$, the fraction in (10a) remains positive and ϑ_2 increases steadily. Consequently, ϑ_2, remaining in the first quadrant, increases to $\arctan(\gamma/\alpha)$. As ϑ_1 increases from $\frac{1}{2}\pi$, the fraction in (10a) remains positive until ϑ_1 has increased to $\arctan(-\delta/\alpha)$, whereupon ϑ_2, remaining in the first quadrant, has increased to $\frac{1}{2}\pi$. As ϑ_1 further increases to $\arctan(-\beta/\gamma)$ (in the second quadrant), ϑ_2 increases from $\frac{1}{2}\pi$ to π. As ϑ_1 increases from $\arctan(-\beta/\gamma)$ to π, $\tan\vartheta_2$ is positive and ϑ_2 increases steadily from π to $\arctan(\beta/\delta)$ (in the third quadrant). Continuing in this manner, we may trace the course of Ω^+ until it closes on the torus. The course of Ω^- can then easily be followed, as it differs from that of Ω^+ only in that ϑ_2 is increased by π.

From (10), (13), (14) and a consideration of the course of the curves Ω^+, Ω^- on the torus, it follows that

(16) $$\begin{aligned}\sin\vartheta_2 &= \pm M^{-1/2}(\gamma\sin\vartheta_1 + \beta\cos\vartheta_1),\\ \cos\vartheta_2 &= \pm M^{-1/2}(\alpha\sin\vartheta_1 + \delta\cos\vartheta_1),\end{aligned}$$

the $+$ sign holding throughout Γ_1 and the $-$ sign throughout Γ_2. On both curves, we have

(17) $$\begin{aligned}x &= a_1\cos\vartheta_1 + b_1\sin\vartheta_1 + a_2\cos\vartheta_2 + b_2\sin\vartheta_2,\\ y &= c_1\cos\vartheta_1 + d_1\sin\vartheta_1 + c_2\cos\vartheta_2 + d_2\sin\vartheta_2,\end{aligned}$$

(18) $$\begin{aligned}\frac{dx}{d\vartheta_1} &= -a_1\sin\vartheta_1 + b_1\cos\vartheta_1 - (a_2\sin\vartheta_2 - b_2\cos\vartheta_2)\frac{d\vartheta_2}{d\vartheta_1},\\ \frac{dy}{d\vartheta_1} &= -c_1\sin\vartheta_1 + d_1\cos\vartheta_1 - (c_2\sin\vartheta_2 - d_2\cos\vartheta_2)\frac{d\vartheta_2}{d\vartheta_1}.\end{aligned}$$

Substituting from (12) and (16), we have

(19) $$\begin{aligned}x &= a_1\cos\vartheta_1 + b_1\sin\vartheta_1 \pm M^{-1/2}[(a_2\alpha + b_2\gamma)\sin\vartheta_1 + (a_2\delta + b_2\beta)\cos\vartheta_1]\\ y &= c_1\cos\vartheta_1 + d_1\sin\vartheta_1 \pm M^{-1/2}[(c_2\alpha + d_2\gamma)\sin\vartheta_1 + (c_2\delta + d_2\beta)\cos\vartheta_1]\end{aligned}$$

(20) $$\begin{aligned}\frac{dx}{d\vartheta_1} &= (-a_1\sin\vartheta_1 + b_1\cos\vartheta_1)(1 \pm \Delta_1\Delta_2{}^2/M^{3/2}),\\ \frac{dy}{d\vartheta_1} &= (-c_1\sin\vartheta_1 + d_1\cos\vartheta_1)(1 \pm \Delta_1\Delta_2{}^2/M^{3/2}),\end{aligned}$$

where the $+$ sign is to be taken in the case of Γ_1 and the $-$ sign in the case of Γ_2. It is to be noted that the first factors on the right-hand side of equations (20) cannot vanish simultaneously by virtue of (9).

The slope of Γ_1 is given by

$$(23)\qquad \frac{dy}{dx} = \frac{c_1(d_1/c_1 - \tan\vartheta_1)}{a_1(b_1/a_1 - \tan\vartheta_1)},$$

while

$$(24)\qquad \frac{d^2y}{dx^2} = \frac{\Delta_1 M^{3/2}}{(b_1\cos\vartheta_1 - a_1\sin\vartheta_1)^3(M^{3/2} + \Delta_1\Delta_2^{\,2})}.$$

From this it is seen that Γ_1 has a continuously turning tangent which is horizontal when $\vartheta_1 = \arctan(d_1/c_1)$, the two corresponding values of ϑ_1 differing by π, and vertical when $\vartheta_1 = \arctan(b_1/a_1)$, the two values of ϑ_1 again differing by π. dy^2/dx^2 changes sign when and only when $b_1\cos\vartheta_1 - a_1\sin\vartheta_1 = 0$, facts which confirm the convexity of Γ_1.

In the case of Γ_2, dy/dx is formally the same, although now undefined for those values of ϑ_1 for which the factor $(1 - \Delta_1\Delta_2^{\,2}M^{-3/2})$ in $dx/d\vartheta_1$ and $dy/d\vartheta_1$ vanishes. The second derivative, d^2y/dx^2, differs from (24) only in that the second factor in the denominator on the right-hand side is replaced by $(M^{3/2} - \Delta_1\Delta_2^{\,2})$. The behaviour of Γ_2 is then similar to that of Γ_1, save at the points for which

$$(25)\qquad M = \Delta_1^{2/3}\Delta_2^{4/3}.$$

At the corresponding values of ϑ_1, if real, the curve Γ_2 has cusps, as may be confirmed by a consideration of (20). Dividing (25) by $\cos^2\vartheta_1$, and simplifying, we obtain

$$(26)\qquad (\alpha^2 + \gamma^2 - \Delta_1^{2/3}\Delta_2^{4/3})\tan^2\vartheta_1 + 2(\alpha\delta + \beta\gamma)\tan\vartheta_1 + \beta^2 + \delta^2 - \Delta_1^{2/3}\Delta_2^{4/3} = 0,$$

a quadratic in $\tan\vartheta_1$. On computing the discriminant and rearranging the result, we may state

Criterion I. *The curve Γ_2 possesses cusps if and only if*

$$(27)\qquad (\alpha^2 + \beta^2 + \gamma^2 + \delta^2)\Delta_1^{2/3}\Delta_2^{4/3} - (\alpha\beta + \gamma\delta)^2 - \Delta_1^{4/3}\Delta_2^{8/3} > 0.$$

By virtue of the results of Kershner,[6] it follows that the inner boundary, D, of $S_1(+)S_2$ exists if and only if the two ellipses have no real intersections. For this, a convenient test may be derived as follows:

On eliminating the parameters ϑ_1, ϑ_2, we find for S_j the equation

$$S_j:\ (c_j^{\,2} + d_j^{\,2})x^2 - 2(a_jc_j + b_jd_j)xy + (a_j^{\,2} + b_j^{\,2})y^2 = \Delta_j^{\,2}, \qquad (j = 1, 2).$$

[6] *Loc. cit.* I, p. 738.

Multiplying the equation for S_1 by Δ_2^2 and that for S_2 by Δ_1^2 and subtracting, we obtain

$$(28)\qquad \begin{aligned}[&(c_1^2 + d_1^2)\Delta_2^2 - (c_2^2 + d_2^2)\Delta_1^2]x^2 \\ &- 2[(a_1c_1 + b_1d_1)\Delta_2^2 - (a_2c_2 + b_2d_2)\Delta_1^2]xy \\ &+ [(a_1^2 + b_1^2)\Delta_2^2 - (a_2^2 + b_2^2)\Delta_1^2]y^2 = 0,\end{aligned}$$

which is the equation of a conic through the four points of intersection of the two ellipses and through the origin. the equation (28) is of the form

$$(28\text{a})\qquad Ax^2 + 2Bxy + Cy^2 = 0.$$

Considerations of symmetry show that the two ellipses will have real intersections when and only when (28a) can be factored into two linear factors with real coefficients, which will be the case if and only if

$$(29)\qquad B^2 - AC \geqq 0.$$

On substituting for A, B, C their values and simplifying the result, we obtain

CRITERION II. *The inner boundary, D, of $S_1(+)S_2$ will exist if and only if*

$$(30)\qquad \begin{aligned}&(a_1^2 + b_1^2)(c_2^2 + d_2^2) + (a_2^2 + b_2^2)(c_1^2 + d_1^2) \\ &\qquad - 2(a_1c_1 + b_1d_1)(a_2c_2 + b_2d_2) < \Delta_1^2 + \Delta_2^2.\end{aligned}$$

By applying Criteria I and II, we obtain four essentially distinct situations with respect to the curves Γ_2 and D, viz.,

(i) Γ_2 has cusps, D does not exist.

(ii) Γ_2 has cusps, D exists.

(iii) Γ_2 has no cusps, D exists.

(iv) Γ_2 has no cusps, D does not exist.

As pointed out in the general case by van Kampen and Wintner, D, if it exists, has corners only in case (ii). Case (iv) arises only if Γ_2 reduces to a single point. This will occur in the addition of two identical ellipses. It may be mentioned that if the ellipses S_1 and S_2 are coaxial, Criteria I and II assume a much simplified form.

The curves Γ_1 and Γ_2, which have been described above, form the boundaries of connected open subsets of R, as may be seen from (4). The determination of the integer m for each of the subsets, cf. (6), though in general not easy, is straightforward in the case of the vector addition of ellipses. For convenience in description, we suppose one ellipse, which we shall designate

as the first ellipse, fixed with its axes coinciding with the coördinate axes (though this is not essential). The vector sum of the two ellipses will then consist of all the points which fall on the circumference of the second ellipse as the latter is translated so that its centre makes a complete circuit of the first ellipse. The second ellipse is divided by its highest and lowest points into two parts, which we designate as the left-hand and the right-hand halves of the ellipse. Then a point of the vector sum sufficiently near the outer boundary, C, is obtained in two and only two ways. For example, if the point is in the upper half-plane, it may be obtained only when the centre of the second ellipse lies on the upper half of the first ellipse and then in just two ways, once when the point lies on the left-hand half of the second ellipse and once when it lies on the right-hand half. Here m must be 2. If, however, the area swept over by the second ellipse when its centre lies on the upper half of the first ellipse overlaps that swept over by the second ellipse when its centre lies on the lower half of the first ellipse, a situation which will arise only in cases (i) and (ii) above, a point in the overlapping region may be obtained in four ways, twice when the second ellipse has its centre on the upper half of the first and twice when the second ellipse has its centre on the lower half of the first. Hence in those regions interior to Γ_2 but exterior to D, if D exists, m will be 4. In case (i) there is just one such region; in case (ii), there are two.

Finally, it may be noted that Γ_2 is an irreducible algebraic equation of fourth degree and can have at most three double points. As, however, the curve is not rational, its maximum number of double points is two, and this number actually occurs in case (ii). A single double point is possible only when located at the origin, and this situation arises as the limiting case between (i) and (ii). Otherwise, Γ_2 possesses no double points.

THE LINCOLN UNIVERSITY,
CHESTER COUNTY, PENNSYLVANIA.

ON DIRICHLET SERIES INVOLVING RANDOM COEFFICIENTS.* [1]

By PHILIP HARTMAN.

Introduction. With any ordinary [2] Dirichlet series

$$\sum_{n=1}^{\infty} a_n/n^s, \qquad s = \sigma + it, \tag{1}$$

one can associate three numbers C, U, A, which have the following significance: the series (1) is convergent (uniformly convergent, absolutely convergent) for all $s = \sigma + it$ in the half-plane $\sigma \geqq C + \epsilon(U + \epsilon, A + \epsilon)$ but is not convergent (uniformly convergent, absolutely convergent in the half-plane $\sigma \geqq C - \epsilon(U - \epsilon, A - \epsilon)$, where $\epsilon > 0$ is arbitrary. The numbers C, U, A are called the abscissa of convergence, abscissa of uniform convergence, abscissa of absolute convergence, respectively. These numbers may be infinite; however, if one is finite, then all are, and they necessarily satisfy the relations [3]

$$1 \geqq A - C \geqq U - C \geqq 0 \tag{2}$$

$$\tfrac{1}{2} \geqq A - U \geqq 0. \tag{3}$$

It is easy to select sequences $\{a_n\}$ such that one of the relations $A - C = 0$, $U - C = 1$ holds. It was first shown by Toeplitz [4] that $A - U > 0$, and even $A - U > \frac{1}{4} - \epsilon$, is possible. Bohnenblust and Hille [5] have constructed series (1) for which $A - U = \frac{1}{2}$ by generalizing the results of Littlewood [5] on infinite bilinear forms to m-forms. These results on $A - U$ are all based on the correspondence between ordinary Dirichlet series and power series in infinitely many variables established by Bohr.[6]

Instead of considering the series (1), the set [7] of Dirichlet series

$$\sum \pm a_n/n^s \tag{4}$$

* Received January 14, 1939.

[1] Presented to the American Mathematical Society, April 8, 1939.

[2] The methods used in this paper are applicable to more general types of Dirichlet series. In general, the abscissa of uniform convergence must be replaced by the abscissa of boundedness, however.

[3] The relation (2) is trivial. The relation (3) was proved by Bohr [3] with his theory of power series in infinitely many variables. It has been shown by Hardy [5 bis] that (3) may also be proved without this apparatus.

[4] Toeplitz [13].

[5] Bohnenblust and Hille [5]; Littlewood [9 bis].

[6] Bohr [3].

[7] All the results in this paper are equally valid if one replaces (4) by the set of series $\Sigma \exp(i\theta_n) a_n/n^s$, where $\theta_1, \theta_2, \cdots$ are independent variables.

is treated in this paper. It follows from the famous "0 or 1" principle,[8] that the probability,[9] that a given number is an abscissa of convergence (uniform convergence, absolute convergence) for one of the series (4), is either 0 or 1. Let $\bar{C}$, $\bar{U}$, $\bar{A}$ be the three uniquely determined numbers such that the probability, that these numbers are the abscissa of convergence, uniform convergence, absolute convergence, respectively, is 1.

The object of this paper is to investigate the possible bounds for the differences $\bar{A}-\bar{U}$, $\bar{U}-\bar{C}$, $\bar{A}-\bar{C}$. It is easy to show that if these abscissae are finite, then

$$\tfrac{1}{2} \geqq \bar{A}-\bar{C} \geqq \bar{U}-\bar{C} \geqq 0 \tag{5}$$

and that the equalities $\bar{U}-\bar{C}=\frac{1}{2}$, $\bar{A}-\bar{C}=0$ may be attained (section 1).

In view of the difficulty of constructing Dirichlet series (1) for which $A-U>0$, and since it is seen from the example of Toeplitz and of its refinements that these difficulties are presented by the actual selection of the signs $\pm 1 = \operatorname{sgn} a_n$ (and not of the order of magnitude of the a_n), Professor Wintner suggested the problem of determining whether or not $\bar{A}-\bar{U}>0$ is possible for (4); so that, if $\bar{A}-\bar{U}>0$, the difficulty of the actual selection of the sgn a_n is avoided, it being taken care of by the "0 or 1" principle. Without using explicitly Bohr's theory of power series in infinitely many variables, it will be shown in the present paper that

$$\bar{A}-\bar{U}=\tfrac{1}{2} \tag{6}$$

is possible (section 3). Thus, in particular, a new proof that the difference $A-U$ can be as large as $\frac{1}{2}$ is given. In the proof of (6), some results of Paley and Zygmund[10] concerning the Fourier series of periodic functions are generalized so as to be applicable to certain classes of almost periodic functions (section 2).

1. The differences $\bar{A}-\bar{C}$, $\bar{U}-\bar{C}$. It may be supposed that $\bar{A}$ is finite; and, furthermore, that $\bar{A}>1$, otherwise a_n may be replaced by $a_n/n^{\bar{A}-1}$. Since $\bar{A}>1>0$, the abscissa of absolute convergence is defined by the relations

---

[8] Cf., e. g., Kolmogoroff [8]. For a simple proof of the "0 or 1" principle for the case at hand, see Hartman and Kershner [6], footnote 4, p. 815.

[9] Probability concerning the set of series (4) is meant in the usual sense (cf. Steinhaus [12]). In particular, to say that "the probability, that the set of series (4) has a certain property, is 1" is equivalent to saying that "for every fixed $\theta\,(0 \leqq \theta \leqq 1)$ on a set of measure 1, the series

$$\Sigma \phi_n(\theta)\, a_n/n^s$$

has this property," where $\{\phi_n(\theta)\}$ is the sequence of orthogonal functions of Rademacher [11] defined by the relation $\phi_n(\theta) = \operatorname{sgn} \sin 2\pi 2^n \theta$.

[10] Paley and Zygmund [10].

$$(7) \qquad \sum_{n=1}^{k} |a_n| = O(k^{\bar{A}+\epsilon}) \text{ but } \sum_{n=1}^{k} |a_n| \neq O(k^{\bar{A}-\epsilon}) \text{ for every } \epsilon > 0.$$

Now, for a fixed $s = \sigma + it$, the probability, that the series (4) converge,[11] is either 0 or 1, according as the series

$$(8) \qquad \sum_{n=1}^{k} |a_n|^2/n^{2\sigma}$$

diverges or converges. Hence, $\bar{C}$ is the abscissa of convergence of the series (8). Since $\bar{A} > 1$, it follows from (2) that $\bar{C} > 0$; so that $\bar{C}$ is defined by the relations

$$(9) \qquad \sum_{n=1}^{k} |a_n|^2 = O(k^{2\bar{C}+\epsilon}) \text{ but } \sum_{n=1}^{k} |a_n|^2 \neq O(k^{2\bar{C}-\epsilon}) \text{ for every } \epsilon > 0.$$

From the Schwarz inequality, one has

$$\sum_{n=1}^{k} |a_n| \leqq (\sum_{n=1}^{k} a_n^2)^{\frac{1}{2}} (\sum_{n=1}^{k} 1)^{\frac{1}{2}} = O(k^{\bar{C}+\frac{1}{2}+\epsilon}) \text{ for every } \epsilon > 0.$$

Thus, from (7), it is seen that $\bar{A} \leqq \bar{C} + \frac{1}{2}$ or $\bar{A} - \bar{C} \leqq \frac{1}{2}$. Since $\bar{A} \geqq \bar{U} \geqq \bar{C}$, it follows at once that (5) is always true.

It is easy to see that $\bar{A} = \bar{U} = \bar{C}$ is possible, for consider the sequence $\{a_n\}$ defined by

$$a_n = 1 \text{ if } n \text{ is of the form } 2^k$$
$$a_n = 0 \text{ if } n \text{ is not of the form } 2^k.$$

The corresponding Dirichlet series (4) are $\Sigma \pm 1/2^{ns}$. It is clear that for each of these series, $A = U = C = 0$.

On the other hand $\bar{A} - \bar{C} = \bar{U} - \bar{C} = \frac{1}{2}$ is also possible, for consider the $\{a_n\}$ defined by

$$a_n = 1 \text{ if } n \text{ is a prime}$$
$$a_n = 0 \text{ if } n \text{ is not a prime.}$$

In this case, the set of series (4) is $\Sigma \pm 1/p_n^s$, where p_n denotes the n-th prime number. It follows from (7) that $\bar{A} = 1$ and from (9) that $\bar{C} = \frac{1}{2}$. Also,[12] for each of these series, $U = A$, in virtue of the linear independence of the numbers $\log p_n$ with respect to the rational number field. Hence $\bar{A} - \bar{C} = \bar{U} - \bar{C} = \frac{1}{2}$.

Thus, the following theorem has been proved:

THEOREM 1. *For all sequences $\{a_n\}$, the probability, that the various abscissae of convergence of the Dirichlet series*

---

[11] That the convergence of (8) implies that the probability, that the series (4) converge, is 1 was first proved by Rademacher [11]. For a proof of the complete 0 or 1 statement, cf. Kolmogoroff [7].

[12] Bohr [4], p. 76.

$$\Sigma \pm a_n/n^s$$

satisfy the relations

$$\tfrac{1}{2} \geqq A - C \geqq U - C \geqq 0,$$

is 1. *There exist sequences* $\{a_n\}$ *such that the probability, that one of relations* $U - C = \frac{1}{2}$, $A - C = 0$ *is satisfied, is* 1.

2. Almost periodic functions involving random Fourier coefficients. In preparation for the proof that $\bar{A} - \bar{U} = \frac{1}{2}$ is possible, an auxiliary problem will be considered. Let $\{\Lambda_n\}$ be an arbitrary sequence of real numbers and let $\{c_n\}$ be a sequence of numbers such that

$$\sum_{n=1}^{\infty} | c_n |^2 < \infty.$$

To these sequences, there corresponds a set of almost periodic[13] functions of class B_2, whose Fourier series are

$$\Sigma \pm c_n \exp(i\Lambda_n t). \tag{10}$$

The auxiliary problem will be to determine conditions to be imposed on the sequences $\{\Lambda_n\}$, $\{c_n\}$ in order that the probability, that the Fourier series (10) may belong to a function which is almost periodic in the sense of Bohr (u. a. p.), is 1.

It will be convenient to rewrite the set of series (10) by using the Rademacher orthogonal functions (cf. footnote 8) and to restate the problem: What conditions must be imposed on $\{\Lambda_n\}$ and $\{c_n\}$ in order that, for almost all θ, the B_2 almost periodic function

$$f(t, \theta) \sim \Sigma \phi_n(\theta) c_n \exp(i\Lambda_n t) \tag{11}$$

may be almost periodic in the sense of Bohr (u. a. p.)?

Suppose that the numbers Λ_n have the property that there exists another sequence of real numbers $\{\lambda_n\}$, which are linearly independent with respect to the rational number field and such that every Λ_n is of the form

$$\Lambda_n = R_1{}^n\lambda_1 + \cdots R_j{}^n\lambda_j, \qquad j = j_n = j(n), \qquad (n = 1, 2, \cdots) \tag{12}$$

where the $R_\nu{}^n$, $\nu = 1, \cdots, j_n$, $n = 1, 2, \cdots$ are integers.

Let $K_n(t)$ denote the Fejér kernel

$$K_n(t) = \sum_{|\nu| < n} \left(1 - \frac{|\nu|}{n}\right) \exp(i\nu t) = \frac{1}{n}\left(\frac{\sin nt/2}{\sin t/2}\right)^2, \tag{13}$$

and let $K_{n_1 \ldots n_k}(t)$ denote the generalized[14] Fejér kernel

[13] Besicovitch [1].

[14] Bohr [4], p. 73.

(14) $$K_{n_1 \ldots n_k}(t) = K_{n_1}(\lambda_1 t) \cdots K_{n_k}(\lambda_k t).$$

It is clear from (13) and (14) that

(15) $$0 \leqq K_{n_1 \ldots n_k}(t) \leqq n_1 \cdots n_k.$$

Because of the linear independence of the λ_n and because

$$M_x\{K_n(x)\} = 1,$$

it follows that

(16) $$M_x\{K_{n_1 \ldots n_k}(x)\} = 1,$$

where

$$M_x\{\cdots\} = \lim_{x\to\infty} \frac{1}{2x} \int_{-x}^{x} \cdots dx.$$

Put

(17) $$\sigma_{n_1 \ldots n_k}(t, \theta) = M_x\{f(x, \theta) K_{n_1 \ldots n_k}(x - t)\}$$
$$= \sum_{|\nu_1| < n_1} \cdots \sum_{|\nu_k| < n_k} \pm \left(1 - \frac{|\nu_1|}{n_1}\right) \cdots \left(1 - \frac{|\nu_k|}{n_k}\right) c\left(\sum_{a=1}^{k} \nu_a \lambda_a\right) \exp i\left(\sum_{a=1}^{k} \nu_a \lambda_a\right) t,$$

where

$$c_n = c(\Lambda_n) = c(R_1{}^n \lambda_1 + \cdots + R_j{}^n \lambda_j)$$

and

$$c(\Lambda) = 0 \text{ if } \Lambda \neq \Lambda_n, \qquad (n = 1, 2, \cdots).$$

The following lemma will now be proved.

LEMMA 1. *For every θ on a set of measure* 1,

(18) $$\sigma_{n_1 \ldots n_k}(t, \theta) = 0(\log^{\frac{1}{2}} n_1 \cdots n_k), \text{ as } n_1 n_2 \cdots n_k \to \infty,$$

uniformly in t, $-\infty < t < +\infty$.

Proof.[15] It is a consequence of a lemma[16] of Paley and Zygmund, that if $\lambda > 0$ is sufficiently small, there exists a constant C_λ such that for all x,

$$\int_0^1 \exp(\lambda \,|\, f(x, \theta)\,|^2)\, d\theta < C_\lambda.$$

Hence,

$$\frac{1}{2x} \int_{-x}^{x} dx \int_0^1 \exp(\lambda \,|\, f(x, \theta)\,|^2)\, d\theta < C_\lambda.$$

Since the integrand is positive, the order of integration may be interchanged. An application of Fatou's lemma yields, that for almost all θ,

(19) $$\varliminf_{x\to\infty} \frac{1}{2x} \int_{-x}^{x} \exp(\lambda \,|\, f(x, \theta)\,|^2)\, dx < \infty.$$

---

[15] This proof is adapted from the proof of Lemma 4, Paley and Zygmund [10].
[16] Paley and Zygmund [10], Lemma 3.

The function $\Psi(b)$ which is complementary[17] to $\Phi(a) = \exp(\lambda a^2) - 1$ is $0(b \log^{\frac{1}{2}} b)$, so that by Young's inequality

$$ab \leqq \exp(\lambda a^2) - 1 + 0(b \log^{\frac{1}{2}} b),$$

for any $a \geqq 0, b \geqq 2$. Hence

$$\begin{aligned} |f(x,\theta)|\, K_{n_1\ldots n_k}(x-t) &\leqq \exp(\lambda\, |f(x,\theta)|^2) \\ &\quad + 0(K_{n_1\ldots n_k}(x-t) \log^{\frac{1}{2}} K_{n_1\ldots n_k}(x-t)). \end{aligned}$$

Now, if $\lambda > 0$ is sufficiently small, it follows from (15), (16), (17) that

$$\sigma_{n_1\ldots n_k}(t,\theta) \leqq \overline{\lim}\, \frac{1}{2x} \int_{-x}^{x} \exp(\lambda\, |f(x,\theta)|^2)\, dx + 0(\log^{\frac{1}{2}} n_1 \cdots n_k)$$

uniformly in t, $-\infty < t < +\infty$. In virtue of (19), this concludes the proof of Lemma 1.

A solution to the auxiliary problem which will be useful in the next section is given by the following

LEMMA 2. *If the sequence $\{\Lambda_n\}$ is such that for a sequence of linearly independent numbers $\{\lambda_n\}$*

$$(12) \qquad \Lambda_n = R_1{}^n \lambda_1 + \cdots + R_j{}^n \lambda_j, \qquad j = j_n = j(n), \qquad (n = 1, 2, \cdots),$$

where the $R_\nu{}^n$ are integers such that for a fixed integer N

$$(20) \qquad \sum_{\nu=1}^{j} |R_\nu{}^n| \leqq N, \qquad (n = 1, 2, \cdots),$$

then the probability that

$$\Sigma \pm c(\Lambda_n) \exp(i\Lambda_n t)$$

is the Fourier series of a u. a. p. function, is 1, *if for some $\epsilon > 0$,*

$$(21) \qquad \sum_{n=1}^{\infty} |c(\Lambda_n)|^2 j_n \log^{2+\epsilon} j_n < \infty.$$

Proof. In order to show that for a fixed θ, $f(t,\theta)$ may be taken to be u. a. p., it is sufficient[18] to show that the sequence of functions $\{\sigma_{n_1\ldots n_k}(t,\theta)\}$ is uniformly convergent in t, as k and $n_1 = n_2 = \cdots = n_k$ tend to infinity.

Ignoring permutations and those $R_\nu{}^n = 0$, there are only a finite number of different sets $(R_1{}^n, \cdots, R_{j_n}{}^n)$ because of (20). It may be supposed that all of the sets are the same. For if the lemma is proved in this case, the general lemma is an immediate consequence, since the sum of a finite number of u. a. p. functions is u. a. p.

[17] Cf., e. g., Zygmund [14], p. 64.

[18] Bohr [4], pp. 69-75.

Let

$$(22)\qquad Q = \max |R_\nu^n| + 1 \text{ for } \nu = 1, \cdots, j_n.$$

Then, for $m \geqq Q, n_1 = \cdots = n_k = m$,

$$\begin{aligned}\sigma_{n_1 \ldots n_k}(t,\theta) &= \prod_{\nu=1}^{j}\left(1 - \frac{|R_\nu^n|}{m}\right) \sum_{|\nu_1|<m} \cdots \sum_{|\nu_k|<m} \pm c\left(\sum_{a=1}^{k} \nu_a\lambda_a\right) \exp i\left(\sum_{a=1}^{k} \nu_a\lambda_a\right)t \\ &= \prod_{\nu=1}^{j}\left(1 - \frac{|R_\nu^n|}{m}\right) \sum_{|\nu_1|<Q} \cdots \sum_{|\nu^k|<Q} \pm c\left(\sum_{a=1}^{k} \nu_a\lambda_a\right) \exp i\left(\sum_{a=1}^{k} \nu_a\lambda_a\right)t,\end{aligned}$$

where the factor of the k-fold sum is independent of n because of the supposition regarding the sets $(R_1^n, \cdots, R_j^n)$. Now, put

$$(23)\qquad \sigma^k(t,\theta) = \sum_{|\nu_1|<Q} \cdots \sum_{|\nu^k|<Q} \pm c\left(\sum_{a=1}^{k} \nu_a\lambda_a\right) \exp i\left(\sum_{a=1}^{k} \nu_a\lambda_a\right)t,$$

so that

$$(24)\quad \sigma_{n_1 \ldots n_k}(t,\theta) = \prod_{\nu=1}^{j}\left(1 - \frac{|R_\nu^n|}{m}\right)\sigma^k(t,\theta), \quad n_1 = n_2 = \cdots = n_k = m \geqq Q.$$

Thus, the problem of the uniform convergence of the sequence $\{\sigma_{n_1 \ldots n_k}(t,\theta)\}$, $n_1 = \cdots = n_k = m$, as $k, m \to \infty$, is reduced to that of the sequence $\{\sigma^k(t,\theta)\}$, as $k \to \infty$.

Let

$$\Sigma_m \pm c(\Lambda_n) \exp(i\Lambda_n t)$$

denote the sum over the finite number of terms of the series (11) for which $j_n = m$. Thus,

$$\sigma^k(t,\theta) = \sum_{m=1}^{k} (\Sigma_m \pm c(\Lambda_n) \exp(i\Lambda_n t)),$$

so that the uniform convergence of the series

$$(25)\qquad \sum_{m=1}^{\infty} (\Sigma_m \pm c(\Lambda_n) \exp(i\Lambda_n t))$$

must be considered.

Put

$$b(\Lambda_n) = c(\Lambda_n) j_n^{\frac{1}{2}} \log^{(2+\epsilon)/2} j_n.$$

Thus, by (21),

$$(26)\qquad \sum_{n=1}^{\infty} |b(\Lambda_n)|^2 < \infty;$$

also,

$$\sum_{m=1}^{k} (\Sigma_m \pm b(\Lambda_n) \exp(i\Lambda_n t))$$

bears the same relationship to

$$\Sigma \pm b_n \exp(i\Lambda_n t)$$

as $\sigma^k(t,\theta)$ does to (11). Hence, by (26), (24), and Lemma 1,

$$\sum_{m=1}^{k} (m^{\frac{1}{2}} \log^{(2+\epsilon)/2} m \, \Sigma_m \pm c(\Lambda_n) \exp(i\Lambda_n t)) = 0(\log^{\frac{1}{2}} Q \cdots Q) = 0(k^{\frac{1}{2}}),$$

uniformly in t, $-\infty < t < +\infty$, for almost all θ. From which one can conclude the uniform convergence of (25) for almost all θ.

3. The difference $A - U$. In order to prove that there exist sequences $\{a_n\}$ for which $\bar{A} - \bar{U} = \frac{1}{2}$, the following theorem will first be proved.

THEOREM 2. *There exist sequences $\{a_n\}$ such that $a_n = 0$ unless n is of the form*

$$n = p_{i_1} \cdots p_{i_N}, \quad i_1 \leqq \cdots \leqq i_N, \qquad N > 0,$$

where N is a fixed integer and p_i denotes the i-th odd prime, for which the probability, that the abscissae of absolute and of uniform convergence of the series

$$\Sigma \pm a_n/n^s$$

satisfy the relation

$$A - U = \tfrac{1}{2} - \frac{1}{2N},$$

is 1.

It may be remarked that it is known [19] that for any Dirichlet series of the form considered, $\frac{1}{2} - \dfrac{1}{2N}$ is the best upper bound for $A - U$.

Proof. Let

$$(27) \qquad \begin{aligned} a_n &= i_N^{-\frac{1}{2}} \text{ if } n = p_{i_1} \cdots p_{i_N},\ i_1 \leqq \cdots \leqq i_N \\ a_n &= 0 \quad \text{ if } n \text{ is not of the form } p_{i_1} \cdots p_{i_N}. \end{aligned}$$

It is clear from (9) that, in this case, $\bar{C} = \frac{1}{2}$.

The set of Dirichlet series (4) corresponding to (27) may be considered as a set of Fourier series (10). It is of the type treated in Lemma 2; here the sequence $\{-\log n\}$ is the sequence $\{\Lambda_n\}$, $\{-\log p_n\}$ corresponds to $\{\lambda_n\}$, and the number N has the same significance in both places. It follows that, in this case, the probability, that the series (4) is the Fourier series of a u. a. p. function, is 1 if $\sigma > \frac{1}{2}$; for here

$$j_n = i_N \text{ if } n = p_{i_1} \cdots p_{i_N},$$

so that, if $\sigma > \frac{1}{2}$,

$$\sum_{n=1}^{\infty} | c_n |^2 j_n \log^{2+\epsilon} j_n = \Sigma p_{i_1}^{-2\sigma} \cdots p_{i_N}^{-2\sigma} \log^{2+\epsilon} i_N < \infty \quad \text{for any } \epsilon > 0.$$

---

[19] Bohnenblust and Hille [5].

In particular, the probability, that the Dirichlet series (4) is uniformly bounded in every half-plane $\sigma \geqq \frac{1}{2} + \epsilon$, is 1. This [20] implies that $\bar{U} = \frac{1}{2}$.

It remains to be proved that $\bar{A} = 1 - \frac{1}{2N}$. If the series (4) is absolutely convergent, the series of absolute values may be rearranged without affecting the convergence. Consider the rearrangement

$$(28) \qquad \sum_{i_1=1}^{\infty} \sum_{i_2=i_1}^{\infty} \cdots \sum_{i_N=i_{N-1}}^{\infty} p_{i_1}{}^{-\sigma} \cdots p_{i_N}{}^{-\sigma} i_N{}^{-\frac{1}{2}}.$$

The inner sum is convergent if and only if $\sigma > \frac{1}{2}$. Now for $\sigma > \frac{1}{2}$,

$$\sum_{i_N=i_{N-1}}^{\infty} p_{i_N}{}^{-\sigma} i_N{}^{-\frac{1}{2}} \geqq \sum_{i_N=i_{N-1}}^{\infty} i_N{}^{-(\frac{1}{2}+\sigma+\epsilon)} \geqq (\sigma - \tfrac{1}{2} + \epsilon)^{-1} i_{N-1}{}^{-(\sigma-\frac{1}{2}+\epsilon)},$$

where $\epsilon > 0$ is arbitrary. Consequently, the inner double sum of (28) is convergent if and only if $2\sigma - \frac{1}{2} > 1$. Continuing in this fashion, it is seen that (28) is divergent unless $N\sigma + \frac{1}{2} - (N-1) > 1$, or $\sigma > 1 - \frac{1}{2N}$. Hence $\bar{A} \geqq 1 - \frac{1}{2N}$. This completes the proof of Theorem 2.

THEOREM 3. *There exist sequences $\{a_n\}$ such that the probability, that the abscissae of absolute and uniform convergence of the series*

$$\Sigma \pm a_n/n^s$$

satisfies the relation

$$A - U = \tfrac{1}{2},$$

is 1.

Proof. Let $a_n{}^N$, $n = 1, 2, \cdots$, for a fixed N, denote the sequence of numbers defined by (27). Consider the series

$$\sum_{N-1}^{\infty} 2^{-N} \sum_{n=1}^{\infty} \pm a_n{}^N/n^s$$

arranged as a Dirichlet series. It is of the form (4). It is clear that for these series $\bar{C} = \bar{U} = \frac{1}{2}$ and $\bar{A} \geqq 1 - \frac{1}{2N}$ for every N. Hence $\bar{A} \geqq 1$. Because of (5), it follows that $\bar{A} - \bar{U} = \frac{1}{2}$.

---

[20] Bohr [2] or Landau [9].

BIBLIOGRAPHY.

[1]. A. S. Besicovitch, *Almost Periodic Functions*, Cambridge (1932).

[2]. H. A. Bohr, "Über die gleichmässige Konvergenz Dirichletscher Reihen," *Journal für die reine und angewandte Mathematik*, vol. 143 (1913), pp. 203-211.

[3]. ———, "Über die Bedeutung der Potenzreihen unendlich vieler Variablen in der Theorie der Dirichletschen Reihen $\Sigma a_n/n^s$," *Göttinger Nachrichten* (1913), pp. 441-488.

[4]. ———, "Fastperiodische Funktionen," *Ergebnisse der Mathematik und ihrer Grenzgebiete*, Bd. 1, Heft 5 (1932).

[5]. H. F. Bohnenblust and E. Hille, "On the absolute convergence of Dirichlet series," *Annals of Mathematics*, vol. 32 (1931), pp. 600-622.

[5 bis]. G. H. Hardy, "The application of Abel's method of summation to Dirichlet series," *Quarterly Journal of Mathematics*, vol. 47 (1916), pp. 176-192.

[6]. P. Hartman and R. Kershner, "The structure of monotone functions," *American Journal of Mathematics*, vol. 59 (1937), pp. 809-822.

[7]. A. Kolmogoroff, "Über die Summen durch den Zufall bestimmter unabhängiger Grössen," *Mathematische Annalen*, vol. 99 (1928), pp. 309-319; vol. 102 (1930), pp. 484-488.

[8]. ———, "Grundbegriffe der Wahrscheinlichkeitsrechnung," *Ergebnisse der Mathematik und ihrer Grenzgebiete*, Bd. 2, Heft 3; in particular, see the appendix.

[9]. E. Landau, "Über die gleichmässige Konvergenz Dirichletscher Reihen," *Mathematische Zeitschrift*, vol. 4 (1921), pp. 317-318.

[9 bis]. J. E. Littlewood, "On bounded bilinear forms," *Quarterly Journal of Mathematics* (Oxford series), vol. 1 (1930), pp. 164-174.

[10]. R. E. A. C. Paley and A. Zygmund, "On some series of functions," *Proceedings of the Cambridge Philosophical Society*, vol. 26 (1930), pp. 337-357.

[11]. H. Rademacher, "Einige Sätze über Reihen von allgemeinen Orthogonalfunktionen," *Mathematische Annalen*, vol. 87 (1922), pp. 112-138.

[12]. H. Steinhaus, "Les probabilités dénombrables et leur rapport à la théorie de la mesure," *Fundamenta Mathematicae*, vol. 4 (1923), pp. 286-310.

[13]. O. Toeplitz, "Über eine bei den Dirichletschen Reihen auftretende Aufgabe aus der Theorie der Potenzreihen von unendlich vielen Veränderlichen," *Göttinger Nachrichten* (1913), pp. 417-432.

[14]. A. Zygmund, *Trigonometrical Series*, Lwów (1935).

QUEENS COLLEGE,
FLUSHING, N. Y.

A LIMIT THEOREM FOR PROBABILITY DISTRIBUTIONS ON LATTICES.*

By E. R. van Kampen and Aurel Wintner.

Introduction. The problem of the random walk on a k-dimensional cubic lattice has been considered by Pólya.[1] He has found, among other results, an asymptotic formula [2] for the probability belonging to a given point of the lattice. Since in Pólya's work each step of the random walk takes place according to a symmetric Bernouilli distribution, it is natural to ask, what will be the situation if one considers arbitrary, instead of merely cubic lattices, and correspondingly allows that the decision of the random walker takes place according to an arbitrary law. It is this question which will be investigated in what follows.

The treatment will be based on a consistent application of the idea of Fourier-Stieltjes transforms which belong to the convolutions of the discontinuous k-dimensional distributions involved.[3] The content of the results will depend on the relation of the least lattice containing the vectorial differences of the points of the spectrum of the distribution function to the least lattice containing this spectrum and the origin; a relation which may conveniently be described in terms of the co-sets of a subgroup of the group which generates the larger lattice.

The result of Pólya referred to above will be further refined by a statement which describes the asymptotic behavior of the probability distributions on the lattice in terms of a k-dimensional normal distribution. This result will imply a statement as to a uniform dissipation of the probability distribution over the lattice and may be considered as an adaptation of the classical Fourier-Stieltjes deduction of the normal distribution law [4] to the discontinuous case of a lattice.

It is instructive to compare the results with recent considerations of P. Lévy [5] concerning angular variables with discontinuous distributions on a finite number of equidistant points of a circle.

* Received April 17, 1939.

[1] G. Pólya, *Mathematische Annalen*, vol. 84 (1921), pp. 149-160.

[2] *Loc. cit.*,[1] formula (12).

[3] As to terminology and notations, cf. B. Jessen and A. Wintner, *Transactions of the American Mathematical Society*, vol. 38 (1935), § 2-§ 3.

[4] P. Lévy, *Bulletin de la Société Mathématique de France*, vol. 52 (1924), pp. 56-68.

[5] P. Lévy, *Comptes Rendus*, vol. 207 (1938), pp. 444-446.

1. Let $\phi = \phi(E)$ be a distribution function, defined for all Borel sets E of a k-dimensional real Euclidean space $R_x : x = (x^1, \cdots, x^k)$; so that $\phi(E)$ is a non-negative completely additive set function satisfying $\phi(R_x) = 1$. The Fourier-Stieltjes transform of ϕ is defined on a k-dimensional real Euclidean space $R_u : u = (u^1, \cdots, u^k)$ by

$$\Lambda(u;\phi) = \int_{R_x} \exp(iu \cdot x)\phi(dR_x), \tag{1}$$

where the period denotes scalar multiplication. It will be assumed that the scalar second moment

$$\mu_2(\phi) = \int_{R_x} x^2\phi(dR_x), \qquad (x^2 = x \cdot x), \tag{2}$$

is neither 0 nor $+\infty$, and that the origin of R_x is the center of mass of the distribution represented by ϕ i. e., that

$$\int_{R_x} x\phi(dR_x) = 0 \quad (= \text{zero vector}). \tag{3}$$

By the spectrum of ϕ is meant the set S of those points x of R_x for which $\phi(E_x) \neq 0$ holds for every open set E_x containing x. It will be assumed that the spectrum S of ϕ is contained in a lattice $L = L(a_1, \cdots, a_k)$ of R_x. It is understood that $a_1, \cdots, a_k$ denote k linearly independent vectors in R_x, and that $L = L(a_1, \cdots, a_k)$ consists of all points $P_n = a_1 n^1 + \cdots + n^k a_k$ of R_x, where $n = (n^1, \cdots, n^k)$ has arbitrary integral components n^j. Notice that if a point Q of R_x is considered as a Borel set, then only the points Q which are identical with a P_n, but not necessarily all points $Q = P_n$, satisfy the inequality $\phi(Q) \neq 0$.

It will always be assumed that L denotes the least lattice containing S and the origin P_0 of R_x. This involves no loss of generality, since one can replace the k-dimensional R_x, when necessary, by a lower-dimensional R_x.

2. By the vectoral difference A (—) B of two point sets A, B contained in R_x will be meant the set of all those points of R_x which may be written in at least one way in the form $a - b$, where a, b are points of A, B, respectively. Since S is contained in the lattice L, it is clear that L contains S (—) S. Let L_0 denote the least lattice containing S (—) S. This sublattice L_0 of the k-dimensional lattice L is again k-dimensional. For if there existed through the origin P_0 of R_x a hyperplane containing L_0, a parallel hyperplane would contain S. But this contradicts (3), since it was assumed that S is not contained in any hyperplane through the origin P_0 of R_x.

Let Γ denote the group of translations of R_x which defines the lattice L,

and let Γ_0 be the invariant subgroup of Γ which corresponds to the sublattice L_0 of L. Since L is the least lattice containing P_0 and S, it is readily seen that the factor group Γ/Γ_0 is cyclic. In fact, all translations of Γ which correspond to points of S are contained in one co-set of Γ_0, and this co-set is a generator of Γ/Γ_0.

Let Γ_h denote the h-th power of the co-set just mentioned, and r the order of the group Γ/Γ_0, finally L_h the set of those points of the lattice L which correspond to the translations contained in Γ_h. Thus, L_p and L_q are identical or disjoint according as $p \equiv q \pmod r$ is or is not satisfied; and L_1 contains S (also when $r = 1$, in which case $L_0 = L_1 = L$).

3. If S_m denotes the spectrum of the distribution function ϕ_m which is defined for $m = 1, 2, 3, \cdots$ as the convolution $\phi_m(E) = \phi_{m-1}(E) * \phi(E)$, where $\phi_1 = \phi$, then S_m is contained in the set L_m just defined, since $S_m = S_{m-1} (+) S$ by the addition rule of point spectra.[3]

While the set L_m depends only on the residue of $m \pmod r$, the same does not, in general, hold for the set S_m contained in L_m. It will however be shown that the spectrum shows with increasing m a tendency to become periodic, with period r, with respect to m. In fact, if m becomes infinite in such a way as to be restricted to one of the r residue classes $(\text{mod } r)$, then S_m becomes L_m:

$$\lim_{m\to\infty} S_{mr+h} = L_h, \qquad (h = 0, 1, \cdots, r-1).$$

This relation will appear as a consequence of a sharper result which concerns the distribution functions $\phi_m(E)$ themselves, instead of their spectra S_m only. In fact, the result will be to the effect that for large m the probability represented by ϕ_{rm+h} becomes equally distributed over a portion of L_h which exhausts L_h in the limit.

The precise formulation of the result to be proved is

$$(4) \quad \lim_{m\to\infty} m^{\frac{1}{2}k}\phi_{mr+h}(P_n) = (2\pi)^{-\frac{1}{2}k} r^{1-\frac{1}{2}k} (\det \| a_1, \cdots, a_k \|)(\det \| \mu^{jl}(\phi) \|)^{-\frac{1}{2}},$$

where the subscript n of P_n is kept fixed, and $h = h(n)$ denotes the subscript of that of the r lattices L_h which contains P_n; while the elements $\mu^{jl}(\phi)$ of the second matrix occurring in (4) are defined as the momenta of the second order

$$\mu^{jl}(\phi) = \int_{R_x} x^j x^l \phi(dR_x); \qquad (j, l = 1, \cdots, k).$$

Since the least lattice L containing S is k-dimensional, it is clear that the non-negative definite quadratic form

(4 bis) $$F(u) = \sum_{j=1}^{k} \sum_{l=1}^{k} \mu^{jl}(\phi) u^j u^l, \qquad \left(\mu^{jl}(\phi) = \int_{R_x} x^j x^l \phi(dR_x)\right),$$
is positive definite.

3 bis. It will turn out that the m-limit process (4) is uniform in n so long $n = o(m^{\frac{1}{2}})$, where $m \to \infty$. This implies that, as $m \to \infty$, the probability distribution represented by ϕ_{rm+h} on L_h becomes equally distributed on any portion of L_h which is contained in a sphere of radius $o(m^{\frac{1}{2}})$ about the origin P_0 of R_x. It will, therefore, follow from the positivity of the limit (4), that the probability distribution ϕ_{rm+h} on L_h does or does not become equally distributed on that portion of L_h which is contained in a varying sphere about the origin P_0 of R_x according as the value of the set function ϕ for this sphere does or does not tend to 0 as $m \to \infty$.

Actually, it will be shown that the variable region may be chosen in a natural manner in such a way that the k-dimensional normal distribution, with (4 bis) as exponent, appears as the limit distribution, when $m \to \infty$. This result will be formulated in such a fashion as to imply the previous statements, as well as (4) itself.

4. It is clear that the problem concerns a distribution question on the lattice L, and not one on the whole space R_x. Correspondingly, nothing is lost by replacing $L = L(a_1, \cdots, a_k)$ by another lattice $L = L(\bar{a}_1, \cdots, \bar{a}_k)$, if each of the k fundamental vectors $\bar{a}_j$ is obtained from the respective a_j by a homogeneous linear transformation whose matrix A is independent of j and has a non-vanishing determinant.

In view of the adaptation theorem of Minkowski, one can choose the basic vectors a_j of the lattice $L = L(a_1, \cdots, a_k)$ in such a way that the sublattice L_0 has the basis $(ra_1, a_2, \cdots, a_k)$, i. e., that $L_0 = L_0(ra_1, a_2, \cdots, a_k)$. On applying to these a_j the transformation A, mentioned before, and choosing A in such a way that a_j becomes the unit vector, e_j, parallel to the x^j-axis of the coördinate system $x : (x^1, \cdots, x^k)$ for $j = 1, \cdots, k$, one can choose

(5_1) $L = L(e_1, e_2, \cdots, e_k)$; $\qquad$ (5_2) $L_0 = L_0(re_1, e_2, \cdots, e_k)$.

Then L_h, where $h = 0, \cdots, r-1$, results from L_0 by a translation of the form qhe_1, where q is an integer relative prime to r.

Since the lattice point P_n belonging to $n = (n^1, \cdots, n^k)$ becomes

(6) $$P_n = n^1 e_1 + \cdots + n^k e_k = n,$$

one can write (1) in the form

(7) $$\Lambda(u; \phi) = \sum_n \exp(iu \cdot n) \phi(P_n).$$

Similarly, (3) becomes

$$\sum_n n\phi(P_n) = 0; \tag{8}$$

while the assumption as to the finiteness of the moment (2) means the convergence of the series

$$\mu_2(\phi) = \sum_n n^2\phi(P_n), \qquad (n^2 = n \cdot n), \tag{9}$$

where

$$\sum_n \phi(P_n) = 1 \text{ and } \phi(P_n) \geqq 0. \tag{10}$$

Finally, the statement (4), to be proved, reduces to

$$\lim_{m\to\infty} m^{\frac{1}{2}k}\phi_{mr+h}(P_n) = (2\pi)^{-\frac{1}{2}k}r^{1-\frac{1}{2}k}(\det \| \mu^{jl}(\phi) \|)^{-\frac{1}{2}}, \tag{11}$$

where $h = h(n)$ is defined as in (4).

5. It is known [6] that if $\psi(E)$ is any distribution on the k-dimensional Euclidean space R_x, and a point x of R_x is considered as a Borel set, then

$$\psi(x) = \lim_{\lambda\to\infty} (2\lambda)^{-k} \int_{Q^\lambda} \Lambda(u;\psi)\exp(-ix\cdot u)\,dR_u, \tag{12}$$

where Q^λ denotes the k-dimensional cube $|u^j| \leqq \lambda$ about the origin of the space R_u of the Fourier-Stieltjes transform of $\psi(E)$. Since ϕ_m was defined as the convolution of ϕ_{m-1} and $\phi_1 = \phi$, application of (12) to the point P_n and the distribution function ϕ_m gives

$$\phi_m(P_n) = \lim_{\lambda\to\infty} (2\lambda)^{-k} \int_{Q^\lambda} \Lambda(u,\phi)^m \exp(-iu\cdot n)\,dR_u, \tag{13}$$

if use is made of (6), (7).

Now, it is clear from (6) and from the last remark of § 2 that the function (7) of $u = (u^1, \cdots, u^k)$ has with respect to each of the k components u^j the primitive period 2π. Consequently, the average (13) reduces to the integral

$$\phi_m(P_n) = (2\pi)^{-k} \int_{Q^\lambda} \Lambda(u;\phi)^m \exp(-iu\cdot n)\,dR_u \tag{14}$$

over the k-dimensional cube $Q^\pi : |u^j| \leqq \pi$.

6. Next, it will be shown that, except at the r points u_h of Q^π which are defined (mod 2π) by

$$u_h = (u_h^1, u_h^2, \cdots, u_h^k) = (2\pi h/r, 0, \cdots, 0); \quad (h = 0, \cdots, r-1), \tag{15}$$

[6] Cf. E. K. Haviland, *American Journal of Mathematics*, vol. 57 (1935), pp. 569-572.

one has $|\Lambda(u;\phi)| < 1$; while $|\Lambda(u;\phi)|$ becomes 1 at every u_h and, more explicitly,

$$\Lambda(u_h;\phi) = \exp(2\pi i q h/r), \qquad (h = 0, \cdots, r-1). \tag{16}$$

where q is the integer defined in § 4.

In order to prove that $|\Lambda(u;\phi)| < 1$ except at the points (15) of Q^π, it is, in view of (10), sufficient to show that if a point u of Q^π is such that $u \cdot (n - n') \equiv 0 \pmod{2\pi}$ for all those $n = (n_1, \cdots, n_k)$ and $n' = (n'_1, \cdots, n'_k)$ for which the points P_n and $P_{n'}$ are in the spectrum S (i. e., $\phi(P_n) > 0$ and $\phi(P_{n'}) > 0$), then u is one of the r points (15). But this is clear from (16), (5_2) and from the definition of L_0 as the least lattice containing S (—) S. Finally, (16) follows by substituting (15) into (7), taking into account (10) and that description of L_1 which is implied by the remark following (5_2); in fact, S is in L_1 by the end of § 2. The same reasoning also shows that (16) may be replaced by the sharper statement

$$\Lambda(u + u_h;\phi) = \Lambda(u;\phi)\exp(2\pi i q h/r), \qquad (h = 0, \cdots, r-1), \tag{17}$$

for every point u of R_u.

6 bis. It may be verified from (15) and the description of L_h in the remark following (5_2), that (14) reduces in virtue of (17) to

$$\phi_{rm+h}(P_n) = (2\pi)^{-k}\, r \int_J \Lambda(u;\phi)^{rm+h} \exp(-iu\cdot n)\, dR_u, \tag{18}$$

where h denotes the subscript of that of the r lattices L_h which contains P_n, and J is the k-dimensional interval

$$J:\quad -\pi/r \leqq u_1 < \pi/r, \quad -\pi \leqq u_j < \pi, \qquad (j = 2, 3, \cdots, k), \tag{19}$$

r being the integer defined in § 2.

Furthermore, if $K(\rho)$ denotes the sphere

$$K:\quad |u| \leqq \rho, \tag{20}$$

then § 6 assures for every sufficiently small $\rho > 0$ the existence of a positive fraction $\theta = \theta_\rho$ which has the property that

$$|\Lambda(u;\phi)| \leqq \theta < 1, \text{ if } u \text{ is in } J - K(\rho), \tag{21}$$

$J - K(\rho)$ denoting the complement of (20) with respect to (19).

7. It will now be easy to carry out the proof of (11) by adapting to the present case the considerations which Pólya [1] has applied in his particular case.

First, it is clear from (21) that if ρ is any sufficiently small fixed positive number, then, as $m \to \infty$,

(22) $m^{\frac{1}{2}k} \int\limits_{J-K(\rho)} \Lambda(u;\phi)^{rm+h} \exp(-iu\cdot n)\,dR_u \to 0$ uniformly for all n.

Hence, it is seen from (18) that (11) will be proved if one shows that, for a sufficiently small $\rho > 0$,

$$\int\limits_{K(m^{\frac{1}{2}}\rho)} \Lambda(m^{-\frac{1}{2}}u;\phi)^{rm+h} \exp(-im^{-\frac{1}{2}}u\cdot n)\,dR_u \to (2\pi)^{\frac{1}{2}k} r^{-\frac{1}{2}k} (\det \| \mu^{jl}(\phi) \|)^{-\frac{1}{2}},$$

(23) where $m \to \infty$, holds uniformly for $|n| <$ const.,

if const. is arbitrarily fixed.

It is clear from (3), (4 bis) and from the assumption that (2) is finite, that (1) is for small $|u|$ of the form

$$\Lambda(u;\phi) = 1 - \tfrac{1}{2}F(u) + o(u^2), \qquad (|u| \to 0).$$

Hence, as $|u| \to 0$,

(24) $\Lambda(m^{-\frac{1}{2}}u;\phi) = 1 - \frac{1}{2}F(u)/m + o(u^2)/m$ uniformly in m.

Since the form (4 bis) is positive definite, one sees from (24) that if $\rho > 0$ is sufficiently small,

(25) $|\Lambda(m^{-\frac{1}{2}}u;\phi)|^{rm+h} \leqq \exp(-\frac{1}{4}rF(u))$ for $|u| \leqq m^{\frac{1}{2}}\rho$.

For a given $\epsilon > 0$, choose a $\lambda = \lambda(\epsilon) > 0$ so large that

(26) $$\int\limits_{R_u-K(\lambda)} \exp(-\tfrac{1}{4}rF(u))\,dR_u < \epsilon, \qquad (\lambda = \lambda(\epsilon)),$$

where $R_u - K(\rho)$ denotes the complement of (20) with respect to the whole u-space. Then, by (25),

(27) $$\Big| \int\limits_{K(m^{\frac{1}{2}}\rho)-K(\lambda)} \Lambda(m^{-\frac{1}{2}}u;\phi)^{rm+h} \exp(-im^{-\frac{1}{2}}u\cdot n)\,dR_u \Big| < \epsilon$$

for $m > \lambda^2/\rho^2$, where $\lambda = \lambda(\epsilon)$ is independent of n.

On the other hand, (24) implies that, as $m \to \infty$,

(28) $\Lambda(m^{-\frac{1}{2}}u;\phi)^{rm+h} \to \exp(-\frac{1}{2}rF(u))$ uniformly for $|u| <$ Const.,

where Const. is arbitrary. Choosing Const. $= \lambda = \lambda(\epsilon)$, one sees from (28) that, as $m \to \infty$,

(29) $$\int\limits_{K(\lambda)} \Lambda(m^{-\frac{1}{2}}u;\phi)^{rm+h} \exp(-im^{-\frac{1}{2}}u\cdot n)\,dR_u \to \int\limits_{K(\lambda)} \exp(-\tfrac{1}{2}rF(u))\,dR_u$$

for every positive λ and for every n. It follows, therefore, from (26) and (27) that, as $m \to \infty$,

$$(30)\quad \int\limits_{K(m^{\frac{1}{2}}\rho)} \Lambda(m^{-\frac{1}{2}}u;\phi)^{rm+h}\exp(-im^{-\frac{1}{2}}u\cdot n)\,dR_u \to \int\limits_{R_u} \exp(-\tfrac{1}{2}rF(u))\,dR_u$$

for every n. Now, on rotating the positive definite quadratic form (4 bis) into its diagonal form, one sees that (30) implies the statement (23).

This completes the proof of (11), hence also that of (4).

8. Next, the refinement of (4) mentioned at the end of § 3 bis will be established. In order to formulate this statement in a precise form, let $G(x)$ denote the reciprocal quadratic form belonging to the positive definite form (4 bis); so that

$$(31)\quad G(x) = \sum_{j=1}^{k}\sum_{l=1}^{k} \nu^{jl}(\phi)x^j x^l, \text{ where } \| \nu^{jl}(\phi) \| \, \| \mu^{jl}(\phi) \| = \| \delta_{jl} \|.$$

It will be shown that if $h = h(n)$ denotes the subscript of that sublattice L_h which contains P_n, and n varies in such a way that, for a fixed point x of R_x,

$$(32)\qquad n = m^{\frac{1}{2}}x + o(m^{\frac{1}{2}}) \text{ as } m \to \infty,$$

then, as $m \to \infty$,

$$(33)\qquad m^{\frac{1}{2}k}\phi_{rm+h}(P_n) \to C\exp(-\tfrac{1}{2}G(x)/r),$$

where C denotes the positive constant

$$(34)\qquad C = (2\pi)^{-\frac{1}{2}k}r^{1-\frac{1}{2}k}(\det \| \mu^{jl}(\phi) \|)^{-\frac{1}{2}}.$$

According to (31), the limit on the right of (33) is proportional to the density of a k-dimensional normal distribution. Since the constant (34) is identical with the limit (11) established by § 7, the statement (33), to be proved, is a generalization of the relation (11), which follows by choosing $x = 0$ in (32).

In order to prove that (32) implies (33) for every fixed x, it is, in view of (18) and (22), sufficient to show that if (32) is satisfied, then, for a sufficiently small $\rho > 0$,

$$(35)\quad \int\limits_{K(m^{\frac{1}{2}}\rho)} \Lambda(m^{-\frac{1}{2}}u;\phi)^{rm+h}\exp(-im^{-\frac{1}{2}}u\cdot n)\,dR_u \to (2\pi)^k C/r\cdot\exp(-\tfrac{1}{2}G(x)/r),$$

uniformly for those n which satisfy (32) uniformly for a fixed x.

On rotating the positive definite quadratic form (4 bis) into its diagonal form, one readily finds the standard relation

$$(36)\quad \int\limits_{R_u} \exp(-\tfrac{1}{2}rF(u))\exp(-iu\cdot x)\,dR_u = (2\pi)^k C/r\cdot\exp(-\tfrac{1}{2}G(x)/r).$$

Hence, it is clear from (26) and (27) that the statement (35), to be proved, is equivalent to the statement that if (32) is satisfied, then, for every positive λ,

$$(37)\quad \int_{K(\lambda)} \Lambda(m^{-\frac{1}{2}}u;\phi))^{rm+h} \exp(-im^{-\frac{1}{2}}u\cdot n)\,dR_u \to \int_{K(\lambda)} \exp(-\tfrac{1}{2}rF(u))\exp(-iu\cdot x)\,dR_u.$$

Since (37) is clear from (28) and (32), the proof is complete.

It is seen from the proof that (37) holds uniformly in n if the o-term of (32) is supposed to hold uniformly with respect to n. This clearly proves all of the statements made in § 3 bis.

THE JOHNS HOPKINS UNIVERSITY.

ON A FAMILY OF SYMMETRIC BERNOULLI CONVOLUTIONS.*

By PAUL ERDÖS.

1. For any fixed real number a in the interval $0 < a < 1$, let $\lambda = \lambda(x; a)$, $-\infty < x < +\infty$, denote the distribution function which is defined as the convolution of the infinitely many symmetric Bernoulli distribution functions $\beta(a^{-n}x)$, $-\infty < x < +\infty$, where $n = 0, 1, 2, \cdots$, and $\beta(x)$ denotes the function which is 0, $\frac{1}{2}$ or 1 according as $x < -1$, $|x| \leqq 1$ or $x > 1$. In other words, $\lambda(x; a)$ is the distribution function whose Fourier-Stieltjes transform is the infinite product

$$L(u; a) \equiv \int_{-\infty}^{+\infty} e^{iux} d_x\lambda(x; a) = \prod_{n=0}^{\infty} \cos(a^n u); \qquad -\infty < u < +\infty.$$

It is known[1] that if a value of a is not such as to make the (monotone) function $\lambda(x; a)$ of x absolutely continuous for $-\infty < x < +\infty$, then $\lambda(x; a)$ is purely singular, that is to say such as to have neither a discontinuous nor an absolutely continuous component in its Lebesgue decomposition. It is also known[2] that the set of those points of the x-axis at which the non-decreasing function $\lambda(x; a)$ is increasing either is the interval $-(1-a)^{-1} \leqq x \leqq (1-a)^{-1}$ or a nowhere dense perfect zero set contained in this interval, according as $a \geqq \frac{1}{2}$ or $a < \frac{1}{2}$. While this clearly implies that $\lambda(x; a)$ is purely singular if $a < \frac{1}{2}$, it does not imply that $\lambda(x; a)$ is absolutely continuous if $a \geqq \frac{1}{2}$. On the other hand, it is known[3] that if a has any of the values $\frac{1}{2}, (\frac{1}{2})^{1/2}, (\frac{1}{2})^{1/3}, \cdots$, then $\lambda(x; a)$ is absolutely continuous, and that $\lambda(x; (\frac{1}{2})^{1/k})$ acquires derivatives of arbitrary high order as $k \to \infty$, i.e., as $a = (\frac{1}{2})^{1/k} \to 1$.

2. Thus, one might be inclined to expect that the "smoothness" of $\lambda(x; a)$ for $-\infty < x < +\infty$ cannot decrease when a is increasing, and that $\lambda(x; a)$, being absolutely continuous if $a = \frac{1}{2}$, is absolutely continuous if $\frac{1}{2} < a < 1$, and not only if $a = (\frac{1}{2})^{1/k}$. However, it turns out that such is

* Received May 12, 1939.

[1] B. Jessen and A. Wintner, *Transactions of the American Mathematical Society*, vol. 38 (1935), p. 61.

[2] R. Kershner and A. Wintner, *American Journal of Mathematics*, vol. 57 (1935), pp. 546-547.

[3] A. Wintner, *American Journal of Mathematics*, vol. 57 (1935), p. 837.

not the case. For instance, it will be shown that if a has the "Fibonacci" value $\frac{1}{2}(5^{1/2}-1)$, a value which lies between $\frac{1}{2}$ and $(\frac{1}{2})^{1/2}$, then $\lambda(x;a)$ is not absolutely continuous and is therefore purely singular (though nowhere constant on its range $|x| \leqq (1-\alpha)^{-1}$). A corresponding a-value between $(\frac{1}{2})^{1/2}$ and $(\frac{1}{2})^{1/3}$ is, for instance, the positive root of the cubic equation $a^3+a^2-1=0$. That $\lambda(x;a)$ is singular for these algebraic irrationalities a, will be proved by showing that the necessary condition $L(u;a)\rightarrow 0$, $u\rightarrow\pm\infty$, of the Riemann-Lebesgue lemma is not satisfied at these particular a-values.

Let α be a real algebraic integer which satisfies the inequality $\alpha>1$ and is such that, if m denotes the degree of α, and α_j, where $j=2,\cdots,m$, are the conjugates of α, then $|\alpha_j|<1$ for all j. Since $\alpha^n+\alpha_2{}^n+\cdots+\alpha_m{}^n$ is a rational integer for $n=0,1,2,\cdots$, it is clear that there exists a positive number $\theta<1$ which has the property that the distance between α^n and the nearest integer to α^n is less than θ^n for every n.

Now choose $a=1/\alpha$. Then, since $L(u;a)=\prod_{n=0}^{\infty}\cos(a^n u)$, one has, for every positive integer k,

$$L(\pi\alpha^k;a)=C\prod_{n=1}^{k}\cos(\alpha^n\pi),\text{ where } C=\prod_{n=0}^{\infty}\cos(\alpha^{-n}\pi)$$

is a non-vanishing constant, since every $\cos(\alpha^{-n}\pi)\neq 0$. Consequently by the above definition of the positive number $\theta<1$,

$$|L(\pi\alpha^k;a)|=|C|\prod_{n=1}^{k}|\cos(\alpha^n\pi)|\geqq C'\prod_{n=1}^{k}|\cos(\theta^n\pi)|,$$

where $C'=|C|\,\Pi^*\,|\cos\theta^n\pi|$ and the product Π^* runs through those values of n for which $\theta^n<\frac{1}{2}$. Hence, for every k,

$$|L(\pi\alpha^k;a)|\geqq C'\prod_{n=1}^{\infty}|\cos(\theta^n\pi)|=\text{const.}>0,$$

if θ is chosen to be distinct from $\frac{1}{2}$. Since $\alpha^k\rightarrow+\infty$ as $k\rightarrow+\infty$, it follows that $L(u;a)$ does not tend to 0 as $u\rightarrow\infty$, and so the distribution function $\lambda(x;a)$ is singular, for any $a=1/\alpha$ of the type described above.

It seems to be likely that these a are clustering at $a=1$ (this would imply that these a lie everywhere dense between $a=0$ and $a=1$).

3. Needless to say, $L(u;a)\rightarrow 0$, $u\rightarrow\infty$, only is a necessary condition in order that $\lambda(x;a)$ be absolutely continuous. In fact, it is known[4] that

[4] R. Kershner, *American Journal of Mathematics*, vol. 58 (1936), pp. 450-452.

if a has any rational value which is not the reciprocal value of an integer, then there exists a positive $\gamma = \gamma(a)$ such that $L(u;a) = O(|\log u|^{-\gamma})$ as $u \to \infty$, whether the positive number $a(<1)$ is or is not greater than $\frac{1}{2}$. (It is easy to see[5] that if $a = \frac{1}{3}, \frac{1}{4}, \cdots$, then $L(u;a) \to 0$ does not hold; while $L(u;a) = (\sin u)/u$ if $a = \frac{1}{2}$.) Actually, it may be shown that, whether $a > \frac{1}{2}$ or $a < \frac{1}{2}$, the Fourier-Stieltjes transform $L(u;a)$ tends, as $u \to \infty$, to 0, not only when a is any rational number distinct from $\frac{1}{3}, \frac{1}{4}, \cdots$, but also for all irrational values of a which do not belong to a certain enumerable set.

In order to prove this, notice first that all values a between 0 and 1 which do not belong to a certain enumerable set are known[6] to possess the following property: There does not exist any number $b > 0$ in such a way that if ϵ_n denotes, for fixed a and fixed b, the distance between ba^{-n} and the nearest integer to ba^{-n}, then $\epsilon_n < \frac{1}{2}(a^{-1}+1)^{-2}$ for every sufficiently large n. Let a be chosen such as to possess this property.

Suppose, if possible, that $L(u;a)$ does not tend to 0 as $u \to \infty$, i.e., that there exists a sequence $u_1, \cdots, u_j, \cdots$ for which one has $u_j \to \infty$ as $j \to \infty$, while $|L(u_j;a)| > c$ holds for a sufficiently small positive $c = c(a)$ which is independent of j. Clearly, one can choose these u_j in such a way that the sequence $\{b_j\}$ defined by $b_j = u_j a^k$ tends, as $j \to \infty$, to a limit, say b, where $k = k_j$ denotes the unique positive integer satisfying $a < u_j a^k \leqq 1$; so that $k_j \to \infty$ as $j \to \infty$, and $a \leqq b \leqq 1$. But $|L(u_j;a)| > c$ may be written in the form

$$|L(b_j a^{-k};a)| = |\prod_{n=0}^{\infty} \cos(b_j a^{n-k})| > c > 0$$

for every j and $k = k_j$. Since $k_j \to \infty$ and $b_j \to b$ as $j \to \infty$, it follows by an obvious adaptation of the inequalities applied in § 2, that b has the property excluded above by the choice of a. This contradiction completes the proof of the fact that $L(u;a) \to 0$, $u \to \infty$, holds for any of the a-values under consideration.

THE INSTITUTE FOR ADVANCED STUDY.

---

[5] Cf. B. Jessen and A. Wintner, *loc. cit.*, Example 1.

[6] C. Pisot, *Annali di Pisa*, ser. 2, vol. 7 (1938), p. 238.

ASYMPTOTIC DISTRIBUTIONS AND THE ERGODIC THEOREM.*

By PHILIP HARTMAN and AUREL WINTNER.

Introduction. Since G. D. Birkhoff's ergodic theorem [1] may be formulated and proved, without any modification of his proof, as a fact in abstract measure theory, it became customary to consider his theorem only from this point of view, and not from the point of view of the needs of classical and statistical mechanics, which originated the ergodic theorem. The object of the present paper is to return to the point of view of these actual needs, by first transforming the dynamical content of the ergodic theorem into a statement regarding asymptotic distribution functions,[2] and then refining the resulting statement so as to imply a "quantitative" refinement of the "qualitative" recurrence theorem of Poincaré. The point is that, fortunately, the conditions for this approach to the dynamical content [3] of the ergodic theorem are always satisfied in the applications to classical and statistical mechanics, but not in the case of an abstract model which carries a Lebesgue measure and nothing else. From the point of view of the applications, it may be assumed that the underlying model is given in terms of the following assumptions [4]:

Let V be a compact, locally Euclidean, n-dimensional manifold of class C^1, on which there is defined a vector function $F = F(x)$ of the position x in such a way that F has n components, continuous partial derivatives, and satisfies the incompressibility condition $\operatorname{div} F(x) \equiv 0$ for the general solution of the system of differential equations which are represented by $dx/dt = F(x)$. It is understood that the compactness of V implies, in view of the local

* Received July 10, 1939.

[1] G. D. Birkhoff, "Proof of the ergodic theorem," *Proceedings of the National Academy of Sciences*, vol. 17 (1931), pp. 656-660.

[2] Cf. A. Wintner, "Remarks on the ergodic theorem of Birkhoff," *ibid.*, vol. 18 (1932), pp. 248-251; B. Jessen and A. Wintner, "Distribution functions and the Riemann zeta function," *Transactions of the American Mathematical Society*, vol. 38 (1935), pp. 48-88; P. Hartman, E. R. van Kampen and A. Wintner, "Asymptotic distributions and statistical independence," *American Journal of Mathematics*, vol. 61 (1939), pp. 477-486. In these papers, further references may be found.

[3] It must, however, be said that the dynamical character of the problem is not used to its full extent, since the incompressibility condition, $\operatorname{div} F(x) \equiv 0$, is only a necessary condition for a system $dx/dt = F(x)$ to be a canonical system.

[4] Not all of these assumptions will be used in the sequel. In particular, it is not necessary to start with differential equations; so that purely topological conditions would be sufficient.

existence and uniqueness theorem for the solutions of ordinary differential equations, that every solution path exists for $-\infty < t < +\infty$, and that incompressibility refers to the Euclidean choice of the Lebesgue measure μ on V.

1. Since V is a compact, locally Euclidean manifold, $0 < \mu(V) < +\infty$. Let P denote an arbitrary point of V, and $\tau^t P$ that point of V which represents, at an arbitrary t, the state determined by the initial position P assigned to $t = 0$. Then not only is τ^t, for every fixed t, a topological transformation of V into itself, but, in addition, $\tau^t P$ is continuous in the product space $V \times T$, where T denotes the t-axis, $-\infty < t < +\infty$. Furthermore, by the incompressibility condition one has $\mu(E) = \mu(\tau^t E)$ for every t and for every Borel subset E of V. These are the assumptions that will be used in the sequel.

While Birkhoff's ergodic theorem may be formulated so as to hold if μ is any abstract Lebesgue measure on any abstract space V, the following considerations cannot be carried through unless one postulates for μ an enumerable sequence of basis sets on V, and assumes that the analogue of the Helly theory of monotone functions is valid. In addition, although Birkhoff's theorem holds, according to an observation of Wiener, if only "measurable sets" and not "points" of V are defined, the following considerations depend essentially on the notion of the "path" of a point, and not only on the notion of the "tube" of a set.

2. By a distribution function $\phi = \phi(E)$ on V is meant a non-negative, absolutely additive set function which is defined for every Borel set E in V in such a way that $\phi(V) = 1$. A Borel set E is called a continuity set of an absolutely additive set function ψ if $\psi(E_*) = \psi(E^*)$, where E_* denotes the interior and E^* the closure of E. If, for every value of a parameter $w > 0$, there is given a distribution function ϕ_w on V, then the sheaf of distribution functions ϕ_w, $0 < w < \infty$, is said to tend to a distribution function ϕ, as $w \to \infty$, if $\phi_w(E) \to \phi(E)$ holds for every fixed continuity set E of ϕ. Since V is locally Euclidean, the separation and compactness theorems of the Helly-Radon theory of absolutely additive set functions are applicable. Furthermore, if a sheaf of distribution functions ϕ_w tends, as $w \to \infty$, to an absolutely additive set function ψ on every continuity set of the latter, then the limit function ψ must be a distribution function, since the compactness of V implies that $\psi(V)$ cannot be distinct from 1.

3. The ergodic theorem of G. D. Birkhoff[5] states that if $f(P)$ is a

[5] Because of some of the reproductions of Birkhoff's proof, it has become customary to state that Birkhoff has proved his ergodic theorem only for bounded L-integrable

Lebesgue integrable function of the position P on V, then the time average $M(f(\tau^t P))$, where

$$M(g(t)) = \lim_{v-u\to\infty} \frac{1}{v-u} \int_u^v g(t)\,dt,$$

exists for almost all points P of V. The excluded zero set depends, in general, on the given function f, and will be denoted by $Z(f)$.

By the path $[P]$ generated by a point P of V is meant the set of all points $\tau^t P$, $-\infty < t < +\infty$, of V. For a path $[P]$ belonging to a given point P of V, and for any interval $u \leqq t \leqq v$, let $\phi^P{}_{uv}(E)$ be defined as the product of the reciprocal length $(v-u)^{-1}$ and of the t-measure of those values of t between u and v for which the point $\tau^t P$ of $[P]$ is in the Borel subset E of V. If the distribution function $\phi^P{}_{uv}$ tends, as $v-u\to\infty$, to a limit distribution function in the sense defined in § 2, then the path $[P]$ will be said to possess an asymptotic distribution function, which will then be denoted by ϕ^P.

Birkhoff's ergodic theorem may now be formulated, without any reference to undetermined summable functions f and corresponding variable zero sets $Z = Z(f)$, as follows:

(I) *The asymptotic distribution function ϕ^P of the path $[P]$ exists whenever the point P does not belong to a certain zero set Z_0 on V.*

It is understood that this Z_0, which may be vacuous, is uniquely determined by the system alone.

The proof proceeds as follows: Let $G_1, G_2, \cdots$ be a sequence of open sets on V which form a topological basis[6] of the manifold V (which is obviously separable). Let $g_n(P)$ denote the characteristic function of G_n. Then, by the above definition of the distribution function $\phi^P{}_{uv}$,

$$\phi^P{}_{uv}(G_n) = \frac{1}{v-u} \int_u^v g_n(\tau^t P)\,dt.$$

Hence, if n is fixed, $\phi^P{}_{uv}(G_n)$ tends, as $v-u\to\infty$, to a limit, $M(g_n(\tau^t P))$, provided that P is not in the zero set $Z(g_n)$. Consequently, $\lim\limits_{v-u\to\infty} \phi^P{}_{uv}(E)$ exists for every $E = G_n$, if P does not belong to the set $\sum\limits_{n=1}^{\infty} Z(g_n)$, which is a

functions. But we were unable to see why Birkhoff's proof, *loc. cit.*[1], should not hold for unbounded, as well as bounded, L-integrable functions. Cf., in fact, N. Wiener, *loc. cit.*[7].

[6] By this is meant that every open subset G of V may be represented as set-theoretical sum of a finite or infinite sequence of sets each of which is a G_n.

zero set. Since the G_n form a topological basis for V, it follows from the Helly-Radon theory, that if P is any fixed point not in this zero set, then there exists a non-negative absolutely additive set function ϕ^P such that the limit of $\phi^P{}_{uv}(E)$, as $v-u\rightarrow\infty$, exists and is represented by $\phi^P(E)$ for any Borel set E which is a continuity set of ϕ^P. Finally, the last remark of § 2 assures that ϕ^P is a distribution function. This completes the proof of (I).

3 bis. Conversely, one can derive the ergodic theorem from (I). In fact, if $f(Q)$ is a bounded, continuous function of the position Q on V, then, for any fixed point P,

$$\frac{1}{v-u}\int_u^v f(\tau^t P)\,dt=\int_V f(Q)\phi^P{}_{uv}(dE),$$

where the integral on the right is a Lebesgue-Stieltjes integral over the space V. It follows that if P is a point not in the zero set Z_0 of (I), the expression on the left tends to the limit

$$M(f(\tau^t P))=\int_V f(Q)\phi^P(dE).$$

If f is an arbitrary summable function on V, there exists a bounded, continuous function g such that the integral $\int|f(P)=g(P)|\,dV$ over the space V is arbitrarily small. But it is known [7] that if α is any positive number, then the set of those points P of V for which

$$\underset{u,v}{\text{l. u. b.}}\ \frac{1}{v-u}\int_u^v |f(\tau^t P)-g(\tau^t P)|\,dt>\alpha$$

has a measure which does not exceed $\alpha^{-1}\int_V|f(P)-g(P)|\,dV$. This clearly implies the validity of the ergodic theorem for arbitrary summable functions.

4. For a path $[P]$ generated by a point P in V, let $[P]'$ denote the closure of the point set $[P]$. On the other hand, let $[P]_-$ and $[P]_+$ denote the set of the so-called α- and ω-points [8] of the path of P, respectively. Thus,

[7] N. Wiener, "The ergodic theorem," *Duke Mathematical Journal*, vol. 5 (1939), pp. 1-18; more particularly, Theorem IV. Wiener points out that Birkhoff's proof, *loc. cit.*[1], is based on an estimate of the same type.

[8] J. Hadamard, "Sur les trajéctoires en dynamique," *Journal de Mathématiques*, ser. 3, vol. 3 (1897), pp. 331-387; G. D. Birkhoff, "Quelques théoremes sur le mouvement des systèmes dynamiques," *Bulletin de la Société Mathématique de France*, vol. 40 (1912), pp. 305-323. Hadamard's lemma concerning the "domain" of a motion, when

$[P]_+$ consists of those points Q of V for which one can find a sequence $t_1, t_2, \cdots$ in such a way that $\tau^{t_n}P \to Q$ and $t_n \to +\infty$, as $n \to \infty$; while $[P]_-$ is obtained by replacing the condition $t_n \to +\infty$ by $t_n \to -\infty$. Both sets $[P]_-$, $[P]_+$ are clearly closed. Assuming, as in § 3, that the conditions of § 1 are satisfied, one can state Poincaré's recurrence theorem[9] as follows:

(II) *Both asymptotic closures $[P]_-$, $[P]_+$ of the path $[P]$ are identical with the closure $[P]'$ whenever the point P does not belong to a certain zero set Z^0 on V.*

It is obvious that this recurrence theorem does not imply the ergodic theorem. But it is also clear from the formulation (I) of the latter, that the recurrence theorem is not implied by the ergodic theorem; in fact, if a continuous function of t is not a constant and tends to a limit as $t \to \infty$, it will have an asymptotic distribution function but no recurrence properties. Hence, there arises the question as to a simultaneous refinement of the ergodic and recurrence theorems.

Such a refinement will now be obtained in terms of the spectra of the distribution functions occurring in the formulation (I) of the ergodic theorem. The spectrum of a distribution function $\phi(E)$ is meant to be the set of those points Q of V which have the property that $\phi(G)$ is distinct from zero whenever G is an open set containing Q. Thus, if $\{P\}$ denotes the spectrum of the asymptotic distribution function ϕ^P of the path $[P]$, the points of $\{P\}$ may be characterized as those points of V the immediate vicinity of which is traveled through by $[P]$, not only for certain arbitrarily large t, but also so frequently that the resulting asymptotic probability cannot vanish.

The essential difference between the spectrum $\{P\}$ and the path $[P]$ or its three closures which occur in (II) is illustrated by the following example: In case of the 1-dimensional path $x = \cos w(t)$, where $w = w(t)$ is a continuous function varying with t from $-\infty$ to $+\infty$, not only the path, but also any of its three closures, consists of the interval $-1 \leqq x \leqq +1$; while

---

combined with the theorem (III) below, implies that there exists a zero set Z in V, such that if P and Q are any pair of points in $V - Z$, the spectra $\{P\}$ and $\{Q\}$ (which are always closed sets) cannot have a common point, unless one contains the other. If it would be possible to prove that $\{P\}$ is identical with $\{Q\}$ whenever $[P]$ contains $[Q]$, except for a possible zero set of points P, Q, it would follow that almost all paths are recurrent in the sense of Birkhoff.

[9] Poincaré's proof of this theorem ("Sur le problème des trois corps et les équations de la dynamique," *Acta Mathematica*, vol. 13 (1889), pp. 1-270, more particularly, pp. 67-72) is perfectly legitimate, although, of course, he does not state his result in the concise terminology of Lebesgue's measure theory. The remark that Poincaré's theorem is a result involving Lebesgue's "almost all," was made independently by Van Vleck and by Carathéodory.

the spectrum of the asymptotic distribution function consists of the pair of points $x = \pm 1$, if $w = w(t)$ is suitably chosen. Thus, (I) and (II) together do not imply the following theorem (which, in turn, obviously implies (II)):

(III) *The asymptotic distribution function ϕ^P of the path $[P]$ exists and has the closure $[P]'$ of the path as spectrum $\{P\}$ whenever the point P does not belong to a certain zero set Z on V.*

5. Before proving (III), consider first the relationships of the three uniquely determined zero sets Z_0, Z^0, Z which occur in (I), (II), (III), respectively.

Let $h(t)$, $-\infty < t < +\infty$, be any continuous, nowhere constant, monotone function which satisfies the interpolation conditions $h(n) = n$; $n = 0, \pm 1, \pm 2, \cdots$. Then the function $x = |\cos 2\pi h(t)|$ attains between $t = n$ and $t = n + 1$ every value between $x = 0$ and $x = 1$ exactly twice. Thus, not only is any of the three closures of the 1-dimensional path identical with the interval $0 \leqq x \leqq 1$, but, in addition, the path is recurrent in the sense of Birkhoff.[10] This holds for every $h(t)$. But it is clear that one can choose $h(t)$ such that $x = x(t) = |\cos 2\pi h(t)|$ does not have an asymptotic distribution function.

Assuming that a path of this type is embedded into an incompressible flow τ^t on a V, one sees that a point of the zero set occurring in (I) need not be a point of the zero set in (II). That a point of the zero set of (II) need not be a point of the zero set occurring in (I), may be illustrated by examples of doubly asymptotic paths. Finally, if a motion of the type $x = \cos w(t)$, mentioned in § 4, is thought of as embedded into an incompressible flow τ^t on a V, one sees that a point of the zero set occurring in (III) need not be in either of the zero sets of (I), (II). On the other hand, it is clear from the definitions, that a point in either of the zero sets occurring in (I), (II) is always a point of the zero set in (III).

6. The proof of (III) will be based on the following lemma:

If a fixed Borel set F in V has a positive measure $\mu(F)$ and is such that $E = F$ is a continuity set of the asymptotic distribution function $\phi^P(E)$ for almost all P contained in F, then

$$\int_F \phi^P(F)\, dV_P \neq 0,$$

where F is fixed, dV_P refers to the Lebesgue measure on V, and the integration runs over the points P of the subset F of V.

[10] G. D. Birkhoff, *loc. cit.*[8].

Since $\tau^t P$ is continuous in P and t together, it is clear from the definition (§ 3) of the distribution function $\phi^P{}_{uv}$, that $\phi^P{}_{uv}(E)$ is, for a fixed Borel set E and for arbitrarily fixed u, v, a measurable function of P on V. Hence, if $E \cdot \tau^t E$ denotes the common part of E and of the Borel set $\tau^t E$ into which E is transformed by τ^t, then

$$\int_E \phi^P{}_{uv}(E)\,dV_P = \frac{1}{u-v}\int_u^v \mu(E \cdot \tau^{-t}E)\,dt.$$

This is readily seen from the definition of $\phi^P{}_{uv}$ and from an application of Fubini's theorem to the product space of the P-set E and the t-interval between u and v. But if E is a set F which is a continuity set of ϕ^P for almost all points P of F, then $\phi^P{}_{uv}(F) \to \phi^P(F)$, as $v - u \to \infty$, holds for any such P. Furthermore, the value of $\phi^P{}_{uv}$ is always between 0 and 1. It follows, therefore, from Lebesgue's theorem on term-by-term integration of uniformly bounded sequences, that

$$\int_F \phi^P(F)\,dV_P = \lim_{v-u\to\infty} \frac{1}{v-u}\int_u^v \mu(F \cdot \tau^{-t}F)\,dt.$$

This completes the proof of the lemma, since it is known [11] that the limit on the right cannot be less than $\mu(F)^2/\mu(V)$, while $\mu(F) \neq 0$ by assumption.

7. The lemma, just proved, readily implies that if G is any given open set in V, then $\phi^P(G) \neq 0$ for almost all points P of G.

In fact, if $\phi^P(G) \neq 0$ does not hold for almost all points P of G, there exists in G a Borel set F of positive μ-measure such that $\phi^P(G) = 0$ for every P in F. This set F of positive measure may be chosen to be a closed set. But then $E = F$ is a continuity set of $\phi^P(E)$ for every point P contained in F, since $\phi^P(G) = 0$ for every P contained in F, and $\phi^P(E)$ is non-negative and absolutely additive. Since $\mu(F) > 0$ by assumption, the lemma of § 6 is applicable. Thus,

$$0 < \int_F \phi^P(F)\,dV_P = \int_F 0 \cdot dV_P = 0.$$

This contradiction proves that $\phi^P(G) \neq 0$ for almost all points P of G.

8. It will now be shown that if G is any fixed open subset of V, then the set Z_G of those points P of V for which G contains a point of the path $[P]$, but does not contain a point of the spectrum $\{P\}$, is a zero set.

In order to prove this, let Z^*_G denote the set of those points P of G for

[11] Cf. C. Visser, "On Poincaré's recurrence theorem," *Bulletin of the American Mathematical Society*, vol. 42 (1936), pp. 397-400.

which either the distribution function ϕ^P does not exist or $\phi^P(G)=0$. According to § 7, the set Z^*_G is a zero set. Let $t_1, \cdots, t_k, \cdots$ be a dense sequence of points on the t-axis, and put $Z_G = \sum_{k=1}^{\infty} \tau^{-t_k} Z^*_G$; so that Z_G is a zero set. It is sufficient to show that if a point P is not in Z_G, while a point of the path $[P]$ is in G, then a point of the spectrum $\{P\}$ is in G.

Since $\tau^t P$ is continuous in t, and at least one point of the path $[P]$ is in the open set G, one may select t_k in such a way that $\tau^{t_k} P$ is in G. Since P is not in $Z_G = \sum_{k=1}^{\infty} \tau^{-t_k} Z^*_G$, the point $\tau^{t_k} P$ is in G but not in Z^*_G. Since ϕ^P remains invariant if the transformation τ^t is applied to P, it follows from the definition of Z^*_G that $\phi^P(G)$ is defined and distinct from zero. Since G must contain at least one point of the spectrum $\{P\}$, the proof is complete.

9. The proof of (III), § 4 may now be completed as follows:

If $G_1, \cdots, G_n, \cdots$ is, as in § 3, a topological basis of V, put $Z^* = \sum_{n=1}^{\infty} Z_{G_n}$; so that Z^* is a zero set, by § 8. It will be shown that if P is not in Z^*, then $\{P\}$ is the closure $[P]'$ of the path of P.

Let the point P be selected in such a way that ϕ^P exists, while $\{P\}$ is not the closure of $[P]$. Since $\{P\}$ is closed and is contained in the closure of $[P]$, there exists an open set G which contains points of $[P]$ but no point of $\{P\}$. One obviously may choose this G to be a G_n. Then P is contained in Z_{G_n}; so that P is contained in the zero set $Z^* = \sum_{n=1}^{\infty} Z_{G_n}$. This completes the proof of (III), since the set of points P for which ϕ^P is not defined is also a zero set.

QUEENS COLLEGE,
THE JOHNS HOPKINS UNIVERSITY.

ON THE DISTRIBUTION OF THE VALUES OF REAL ALMOST PERIODIC FUNCTIONS.*

By M. Kac,** E. R. van Kampen and Aurel Wintner.

Introduction. While there is a well developed general theory for the asymptotic distribution functions of a real almost periodic function $f(t)$, practically nothing seems to be known about the more delicate question which concerns the asymptotic density of the roots t of the equation $f(t)=a$, where a is a fixed real number.[1] This question clearly becomes meaningless, unless $f(t)$ is subjected to suitable local restrictions, at least.

The purpose of the present paper is to investigate, with the help of the Kronecker-Weyl approximation theorem, this problem of a-values in case $f(t)$ satisfies suitable conditions, which will be specified later. Actually, the rôle of the Kronecker-Weyl theorem is that of establishing a connection between the problem of a-values and an interesting, though elementary, problem concerning "geometrical probabilities."

1. Suppose that a given set Σ in an n-dimensional Euclidean space Θ: $(x_1,\cdots,x_n)$ satisfies the periodicity condition that if $(x_1,\cdots,x_j,\cdots,x_n)$ is a point of Σ, then so is $(x_1,\cdots,x_j\pm 1,\cdots,x_n)$ for $j=1,\cdots,n$. Let

$$\Theta_0:\qquad (\theta_1,\cdots,\theta_n)\qquad (\mathrm{mod}\ 1) \tag{1}$$

denote the n-dimensional torus obtained from Θ by reduction mod 1, and Σ_0 the set which results from Σ on this reduction. For any direction $\lambda=(\lambda_1,\cdots,\lambda_n)\neq(0,\cdots,0)$ in Θ_0, let Γ_λ be that characteristic

$$\Gamma_\lambda:\qquad \theta_j=\lambda_j t+\phi_j,\qquad (j=1,\cdots,n;\ -\infty<t<+\infty), \tag{2}$$

on Θ_0 which belongs to a given initial phase $(\phi_1,\cdots,\phi_n)$.

It will be assumed that Σ_0 satisfies the following condition of "smoothness": If, for any fixed λ and for any $q>0$, one denotes by $M_\lambda=M_\lambda(\phi_1,\cdots,\phi_n;q)$ the number $(\leqq+\infty)$ of the common points of Σ_0 and of the portion $0\leqq t\leqq q$ of the characteristic Γ_λ, then $M_\lambda(\phi_1,\cdots,\phi_n;q)$ has a finite Lebesgue integral over the n-dimensional phase torus of the ϕ_j.

---

* Received April 11, 1939.

** Fellow of the Parnas Foundation, Lwów (Poland).

[1] We only know a paper by P. Stein, "On the real zeros of a trigonometrical function," *Proceedings of the Cambridge Philosophical Society*, vol. 31 (1935), pp. 455-467, where the superposition $f(t)=\cos t+a\cos\lambda t$ of two incommensurable harmonic vibrations is considered.

Now consider the following problem on geometrical probability: Given the periodic set Σ, what is the probable number $P_\lambda = P_\lambda(l)$ of intersections of Σ with a "needle," of given length l, which is thrown on the Euclidean space Θ in the given direction λ?

It is understood that the spatial probability distribution (mod 1) of the position of the needle is defined such as to be invariant under translations of the coördinate system $(x_1, \cdots, x_n)$. Then, as is well known (Crofton),[2] the probability that the "lower" end point of the needle will lie (mod 1) in an open set of Θ_0 is equal to the Euclidean volume measure of the open set. On placing

$$(3) \qquad | \lambda | = (\lambda_1^2 + \cdots + \lambda_n^2)^{\frac{1}{2}}$$

and noting that, if c is any real number, the segment $c \leqq t \leqq c + q$ of the characteristic (2) has the Euclidean length

$$(4) \qquad l = q \, | \lambda |,$$

one readily verifies from the definition of a Lebesgue integral that if $P_\lambda = P_\lambda(l)$ denotes the probable number in question, then

$$(5) \qquad P_\lambda(l) = \int_{\Theta_0} M_\lambda(\theta_1, \cdots, \theta_n; q) \, d\Theta_0,$$

where $d\Theta_0$ is the Euclidean volume element of (1).

Actually, $P_\lambda(l)$ is proportional to l. In fact, the function $M_\lambda(\phi_1, \cdots, \phi_n; q)$ obviously satisfies the functional equation

$$M_\lambda(\phi_1, \cdots, \phi_n; q) + M_\lambda(\phi_1 + q\lambda_1, \cdots, \phi_n + q\lambda_n; p) = M_\lambda(\phi_1, \cdots, \phi_n; q + p)$$

for arbitrary $q > 0$, $p > 0$. On integrating this identity over the whole phase torus $(\phi_1, \cdots, \phi_n)$, and then using (5) and (4), one sees that

$$P_\lambda(l) + P_\lambda(m) = P_\lambda(l + m), \text{ where } m = p \, | \lambda |.$$

Since $P_\lambda(l)$ is a non-decreasing function of l, it follows that

$$(6) \qquad P_\lambda(l) = l D_\lambda,$$

where D_λ is independent of l. This proves the statement. Clearly, D_λ may be interpreted as the average density of the points of Σ on the needle.

1 bis. It will now be assumed, without loss of generality, that the first component, λ_1, of λ is positive. Then the representation (5) of (6) as an integral over the n-dimensional torus (1) may be reduced to the representation of the average density D_λ as the integral

$$(7) \qquad D_\lambda = \frac{\lambda_1}{| \lambda |} \int_\Phi M_\lambda(0, \phi_2, \cdots, \phi_n; 1/\lambda_1) \, d\Phi \qquad \text{[cf. (3)]}$$

[2] Cf. R. Deltheil, *Probabilités géométriques*, Paris, 1926.

over the $(n-1)$-dimensional torus

$$\Phi:\quad 0 \leqq \phi_j < 1; \qquad (j = 2, \cdots, n). \tag{8}$$

In order to see this, notice first that, on substituting (4) into (6) and then (6) into (5), one obtains

$$D_\lambda = \frac{\lambda_1}{|\lambda|} \int_{\Theta_0} M_\lambda(\theta_1, \theta_2, \cdots, \theta_n; 1/\lambda_1)\, d\Theta_0 \tag{9}$$

if one chooses $q = 1/\lambda_1$. On the other hand, the increase of the first component, $\theta_1 = \lambda_1 t + \phi_1$, of (2) on any t-interval of length $1/\lambda_1$ is 1; so that this increase is precisely the period of the angular function $\theta_1(t) = \lambda_1 t + \phi_1$, which is thought of as reduced mod 1. One can, therefore, verify by a straightforward calculation which is similar to the proof of (6), that the value of the integral

$$\int_\Phi M_\lambda(\phi_1, \cdots, \phi_n; 1/\lambda_1)\, d\Phi \qquad [\text{cf. } (8)] \tag{10}$$

is independent of ϕ_1. Hence, on choosing $\phi_1 = 0$ in (10), one sees from (1) and (8) that (9) is equivalent to the statement (7).

These elementary considerations were only presented in order to supply the "geometrical" background of the problem of a-values, which will be taken up in § 2.

2. Let

$$F = F(\theta_1, \cdots, \theta_n) \tag{11}$$

be a real-valued continuous function of the position on the torus (1). It may be assumed without loss of generality that $F(\theta_1, \cdots, \theta_n)$ actually depends on each of the θ_j. Let $\lambda_1, \cdots, \lambda_n$ be n real numbers which are linearly independent in the rational field; so that

$$f(t) = F(\lambda_1 t + \phi_1, \cdots, \lambda_n t + \phi_n) \tag{12}$$

is, for arbitrary initial phases ϕ_j, an almost periodic function whose modul is generated by the $2\pi\lambda_j$. It may be supposed without loss of generality that

$$\phi_1 = 0 \quad \text{and} \quad \lambda_1 > 0. \tag{13}$$

It will be assumed that the function (11) satisfies, with reference to a fixed real number a, the following condition of "smoothness": If $G_a = G_a(\phi_2, \cdots, \phi_n)$ denotes the number of the roots t of the equation $f(t) = a$ on the interval $0 \leqq t \leqq 1/\lambda_1$, where λ_1 is fixed, then $G_a(\phi_2, \cdots, \phi_n)$ is Riemann integrable (and, therefore,[3] bounded) on the $(n-1)$-dimensional torus (8).

[3] O. Hölder, *Beiträge zur Potentialtheorie*, Stuttgart, 1882 (Tübingen Thesis), p. 3.

For any $T > 0$, let $N_T(a)$ denote the number of the roots t of the equation $f(t) = a$ on the interval $0 \leqq t \leqq T$; so that $N_T(a) : T$ is the density of the a-values of $f(t)$ between $t = 0$ and $t = T$. It follows from the Kronecker-Weyl approximation theorem that the asymptotic density

$$E_\lambda(a) = \lim_{T \to \infty} \frac{N_T(a)}{T} \tag{14}$$

exists for $-\infty < a < +\infty$; and that the value of $E_\lambda(a)$ is the Riemann integral

$$E_\lambda(a) = \lambda_1 \int_\Phi G_a(\phi_2, \cdots, \phi_n) d\Phi \tag{15}$$

over the $(n-1)$-dimensional torus (8).

Let the set Σ_0, defined at the beginning of § 1, be identified with the set of the zeros $(\theta_1, \cdots, \theta_n)$ of the difference $F(\theta_1, \cdots, \theta_n) - a$ on Θ_0; so that

$$\Sigma_0: \qquad F(\theta_1, \cdots, \theta_n) = a, \tag{16}$$

and the function $G_a(\phi_2, \cdots, \phi_n)$ is identical with the function $M_\lambda(0, \phi_2, \cdots, \phi_n; 1/\lambda_1)$, considered in § 1 bis. Then, if the smoothness conditions of both the present § 2 and of § 1 are satisfied, one sees from (7) and (15) that

$$D_\lambda(a) = \frac{E_\lambda(a)}{|\lambda|}. \tag{17}$$

If only the Lebesgue assumptions of § 1 are satisfied by the set (16), then, on applying the ergodic theorem of Birkhoff instead of the approximation theorem of Kronecker-Weyl, one sees that the above considerations remain valid for almost all initial phases $(\phi_1, \cdots, \phi_n)$ of (12), so long as a and the frequencies $\lambda_1, \cdots, \lambda_n$ are fixed.

3. The actual evaluation of the integral (15) is usually difficult even in simple examples. In some cases, this problem of evaluation is facilitated by the following observation:

It is clear from the notations introduced at the beginning of § 1, that if Σ and Σ_0 are thought of as situated in Θ and Θ_0, respectively, then Σ and Σ_0 determine each other uniquely. On the other hand, if both Σ and Σ_0 are thought of as situated in Θ, then Σ_0 determines Σ uniquely, while the converse does not hold; so that two distinct sets Σ_0, say Σ_0 and Σ'_0, may determine the same Σ. Now suppose that the set Σ_0 defined by (16) possesses the property that there exists a Σ'_0 which is a convex hypersurface and belongs in Θ to the same Σ as Σ_0 itself. Then, for the purposes of § 2, one can replace Σ_0 by Σ'_0. And this observation implies that

$$E_\lambda(a) = 2 \,|\, \lambda \,|\, S, \qquad \text{[cf. (3), (14)]}, \tag{18}$$

where S denotes the $(n-1)$-dimensional volume of the orthogonal projection of Σ'_0 on a hyperplane which is perpendicular to the direction $\lambda = (\lambda_1, \cdots, \lambda_n)$.

On applying the rule (18) to

$$F(\theta_1, \theta_2) = \cos\theta_1 + \cos\theta_2, \tag{11 bis}$$

we have found that if λ_1, λ_2 are two incommensurable frequencies and ϕ_1, ϕ_2 two arbitrary phases, then the asymptotic density (14) of the a-values of the almost periodic function

$$f(t) = \cos(\lambda_1 t + \phi_1) + \cos(\lambda_2 t + \phi_2) \tag{12 bis}$$

is given by

$$E_\lambda(a) = 4(|\lambda_1 y| + |\lambda_2 x|) = E_\lambda(-a) \text{ if } 0 \leqq a \leqq 2$$

(and by $E_\lambda(a) = 0$ if $a > 2$), where x, y follow from

$$\cos x = \frac{\{(\lambda_1^2 - \lambda_2^2) + a^2\lambda_1^2\lambda_2^2\}^{\frac{1}{2}} - a\lambda_2^2}{\lambda_1^2 - \lambda_2^2},$$

$$\cos y = \frac{\{\ \}^{\frac{1}{2}} - a\lambda_1^2}{\lambda_2^2 - \lambda_1^2}; \qquad |x| \leqq \pi, \ |y| \leqq \pi.$$

It would be worth while to evaluate, in terms of suitable factors of discontinuity, the density of the a-values of (12) in the more general case

$$F(\theta_1, \cdots, \theta_n) = \sum_{j=1}^{n} r_j \cos\theta_j, \text{ where } r_j = \text{const.} \neq 0.$$

In this case, the assumption of (18) as to the existence of a convex hypersurface is not in general satisfied.

3 bis. On the other hand, this condition is satisfied in the elementary case

$$F(\theta_1, \cdots, \theta_n) = \sum_{j=1}^{n} \beta_j \cdot (\theta_j),$$

where (θ) denotes the least non-negative residue (mod 1) of the real number θ, while the β_j are real constants which, without loss of generality, will be subjected to the normalization $\beta_1^2 + \cdots + \beta_n^2 = 1$. Although (11) is then not continuous on the torus (1), one readily sees that the rule (18) remains applicable. It follows that the asymptotic density (14) of the a-values of

$$f(t) = \sum_{j=1}^{n} \beta_j \cdot (\lambda_j t + \phi_j)$$

exists and may be represented, for arbitrary phases ϕ_j and for arbitrary linearly independent frequencies λ_j, by the product of $|\lambda_{(\beta)}|$ and S_a, where $|\lambda_{(\beta)}|$

denotes the length of the projection $\lambda_{(\beta)}$ of the constant vector $\lambda = (\lambda_1, \cdots, \lambda_n)$ on the line $x_j = \beta_j t$ in the Euclidean $(x_1, \cdots, x_n)$-space, while S_a is the $(n-1)$-dimensional volume of the intersection of the hyperplane $\sum_{j=1}^{n} \beta_j x_j = a$ with the n-dimensional cube

$$\Psi: \qquad 0 \leqq x_j < 1, \qquad\qquad (j = 1, \cdots, n)$$

in a Euclidean $(x_1, \cdots, x_n)$-space.

It may be mentioned that this $(n-1)$-dimensional volume S_a is identical with the density at $t = a$ of the one-dimensional distribution $\tau = \tau(t)$ which one obtains by orthogonal projection of the n-dimensional equidistribution in the cube Ψ on the line $x_j = \beta_j t$, $-\infty < t < +\infty$ of the Euclidean $(x_1, \cdots, x_n)$-space.[4] In other words, S_a is the density at $t = a$ of the convolution of the n one-dimensional distribution functions $\sigma_j(t)$, where σ_j is the equidistribution on the t-interval whose end-points are 0 and β_j.

4. The problem of a-values for almost periodic functions $f(t)$, as considered in § 2 in case $f(t)$ is real-valued, cannot admit of a similar treatment in case $f(t)$ is allowed to be complex-valued for $-\infty < t < +\infty$. For instance, it is clear that if (12) is complex-valued, then the density of the a-values is 0 for almost all a in the complex a-plane, except when the function (11), which may be arbitrarily smooth, has a rather special structure.

There arises, however, a more reasonable question if, in case of a complex-valued $f(t)$, one considers the problem of the distribution of the a-values of $\arg f(t)$, where the almost periodic function $f(t)$ is supposed to be distinct from 0 for every t.

Actually, it will be assumed that $|f(t)| >$ const. > 0 for $-\infty < t < +\infty$. Then, on replacing $f(t)$ by $f(t)/|f(t)|$, one can also assume that the almost periodic function $f(t)$ is such that, for every t,

$$(19) \qquad |f(t)| = 1, \text{ i.e., } f(t) = \exp 2\pi i h(t),$$

where $h(t)$ is real and continuous. And the question concerns the existence (and then the determination) of

$$(20) \qquad \mathrm{E}(\alpha) = \lim_{T\to\infty} \frac{\mathrm{N}_T(\alpha)}{T}, \qquad (0 \leqq \alpha < 1),$$

where $\mathrm{N}_T(\alpha)$ denotes, for a fixed angle α and for a fixed $T > 0$, the number of those points t on the interval $0 \leqq t \leqq T$ at which $((h(t)) = \alpha$, the symbol (h) denoting $h - [h]$.

It will again be assumed that $f(t)$ may be represented in terms of n $(< +\infty)$ initial phases ϕ_j, of n linearly independent real numbers λ_j, and

[4] Cf. A. Sommerfeld, "Eine besonders anschauliche Ableitung des Gaussischen Fehlergesetzes," *Boltzmann Festschrift*, Leipzig (1904), pp. 848-859.

of a continuous function (12) of the position on (1) in the form (11). Then, in virtue of (19),

$$(21) \qquad | F(\theta_1, \cdots, \theta_n) | = 1 \text{ on } \Theta_0.$$

Hence, by a well-known theorem of Bohr, there exist n integers m_j and a continuous real-valued function

$$(22) \qquad H = H(\theta_1, \cdots, \theta_n) \text{ on } \Theta_0$$

in such a way that

$$(23) \qquad F(\theta_1, \cdots, \theta_n) = \exp 2\pi i (\sum_{j=1}^{n} m_j \theta_j + H(\theta_1, \cdots, \theta_n)) \text{ on } \Theta_0.$$

Now, on introducing for the α-values of the function

$$(\sum_{j=1}^{n} m_j \theta_j + H(\theta_1, \cdots, \theta_n)), \text{ where } (u) = u - [u],$$

the assumptions which correspond to the assumptions made in § 2 with regard to the a-values of the real function $F(\theta_1, \cdots, \theta_n)$, one readily sees that the considerations which led to (15) remain valid and lead to a corresponding integral representation of (20).

4 bis. Suppose, in particular, that the mean motion, $2\pi \sum_{j=1}^{n} m_j \lambda_j$, of $f(t)$ is not contained in the modul of the almost periodic function

$$(24) \qquad 2\pi H(\lambda_1 t + \phi_1, \cdots, \lambda_n t + \phi_n) = \arg(f(t) - 2\pi \sum_{j=1}^{n} (\lambda_j t + \phi_j)).$$

Then (11) may be assumed to be of the form

$$(25) \qquad F(\theta_1, \cdots, \theta_n) = \exp 2\pi i (\theta_n + H(\theta_1, \cdots, \theta_{n-1})).$$

Hence, an application of the angular analogue to the rule (15) readily shows that the relative density (20) of the α-values is independent of α.

It should be mentioned that if the function (22), which may depend on all n torus coördinates θ_j, satisfies a Lipschitz condition

$$| H(\theta_1, \cdots, \theta_n) - H(\bar{\theta}_1, \cdots, \bar{\theta}_n) | < C \sum_{j=1}^{n} | \theta_j - \bar{\theta}_j |, \quad (| \theta_j - \bar{\theta}_j | \leqq \tfrac{1}{2}),$$

and if the number $| 2\pi \sum_{j=1}^{n} m_j \lambda_j |$ (i. e., the absolute value of the mean motion of $f(t)$) is sufficiently large with reference to a fixed value of the Lipschitz constant C, then the smoothness conditions alluded to before are automatically satisfied. Thus, application of the analogue to (15) again shows that the asymptotic density (20) of the α-values is independent of α. The constant value of (20) is readily seen to be the absolute value of the mean motion of $f(t)$.

THE JOHNS HOPKINS UNIVERSITY.

A REMARK ON ASYMPTOTIC CURVES.*

By E. R. van Kampen.

In this note a simple geometrical property of asymptotic curves is discussed which apparently has escaped observation up till now. In order to establish the necessary normalizations, a short proof is given for the theorem of Enneper.

Let $\boldsymbol{x} = \boldsymbol{x}(u^1, u^2)$ represent a section of a surface S of negative Gaussian curvature K in 3-dimensional Euclidean x-space. If $\boldsymbol{x}_i$, $\boldsymbol{x}_{ij}$, $i, j = 1, 2$, represent the partial derivatives of $\boldsymbol{x}$, and $\boldsymbol{n}$ denotes the unit vector in the normal direction $\boldsymbol{x}_1 \times \boldsymbol{x}_2$, then the coefficients of the first and second fundamental form of S are $g_{ij} = \boldsymbol{x}_i \cdot \boldsymbol{x}_j$ and $h_{ij} = \boldsymbol{x}_{ij} \cdot \boldsymbol{n}$. A tangent vector of unit length will be represented by $\dot{u}^i$ and also by $\dot{\boldsymbol{x}} = \boldsymbol{x}_i \dot{u}^i$, where $g_{ij}\dot{u}^i\dot{u}^j = 1$, so that the vector is thought of as the tangent vector of some curve on S, which carries the arc length as a parameter. With this convention, the two asymptotic directions, which are real and distinct since $K < 0$, are obtained from the quadratic equation

$$(*) \qquad h_{ij}\dot{u}^i\dot{u}^j = 0 \text{ or } \dot{\boldsymbol{x}} \cdot \dot{\boldsymbol{n}} = 0, \qquad (g_{ij}\dot{u}^i\dot{u}^j = 1).$$

Any curve β of class C^2 on S, tangent to an asymptotic direction, satisfies $\ddot{\boldsymbol{x}} \cdot \boldsymbol{n} = 0$ also, as is seen by differentiation of the identity $\dot{\boldsymbol{x}} \cdot \boldsymbol{n} = 0$. Thus, if the curvature of β is distinct from 0, i. e., $\ddot{\boldsymbol{x}} \neq 0$, the binormal of β is $\pm \boldsymbol{n}$, and the principal normal is $\pm \boldsymbol{n} \times \dot{\boldsymbol{x}}$. The equations of Serret-Frenet now give for the torsion τ of β the equations $\tau = -(\pm \boldsymbol{n} \times \dot{\boldsymbol{x}}) \cdot (\pm \dot{\boldsymbol{n}}) = (\dot{\boldsymbol{x}}\boldsymbol{n}\dot{\boldsymbol{n}})$ and $\tau^2 = \dot{\boldsymbol{n}} \cdot \dot{\boldsymbol{n}}$.

On the other hand, if (g^{ij}) is the inverse matrix of (g_{ij}), then the Cayley-Hamilton equation of the matrix $(g^{ik}h_{kj})$ leads to $Kg_{ij} - 2Hh_{ij} + h_{ik}g^{kl}h_{lj} = 0$, where H is the mean curvature of S. On multiplying the last equation by $\dot{u}^i\dot{u}^j$ and then using $(*)$, one obtains $K + h_{ik}g^{kl}h_{lj}\dot{u}^i\dot{u}^j = 0$, or simply $\dot{\boldsymbol{n}} \cdot \dot{\boldsymbol{n}} = -K$. Thus, if β is tangent to an asymptotic direction at the point P and if the curvature of β is not 0 at P, then the torsion τ of β at P satisfies the equation $\tau^2 = -K$.

The actual value of τ is now determined very easily. In fact, it may be assumed that β has positive geodesic curvature at P. Then its binormal is $\boldsymbol{n}$, and its principal normal is $\boldsymbol{n} \times \dot{\boldsymbol{x}}$. For increasing s, the curve β enters a

* Received April 13, 1939.

region of S, which corresponds to a small positive rotation of the tangent vector $\dot{\boldsymbol{x}} = \boldsymbol{x}_i \dot{u}^i$, and this region of S is on the positive or negative side of the tangent plane of S at P (= osculating plane of β at P) according as a small positive rotation of $\dot{\boldsymbol{x}} = \boldsymbol{x}_i \dot{u}^i$ brings this vector into a position where the signature ϵ of $h_{ij}\dot{u}^i\dot{u}^j$ is $+1$ or -1. It is understood that "positive side" and "negative side" are meant in the sense determined by the orientation of n.

It now follows from a well known determination of the signature of the torsion of a space curve, that the torsion of β has the signature ϵ. Since $\tau = (\dot{\boldsymbol{x}}\boldsymbol{n}\dot{\boldsymbol{n}})$, the direction of $\dot{\boldsymbol{n}}$ is that of $\epsilon\dot{\boldsymbol{x}} \times \boldsymbol{n}$. Thus one obtains the following normalized statement of Enneper's theorem:

I. *If β is a curve on S tangent at P to an asymptotic direction $\dot{\boldsymbol{x}} = \boldsymbol{x}_i \dot{u}^i$ of S, if a small positive rotation of the tangent vector $\dot{\boldsymbol{x}}$ brings this vector into a region where the signature of $h_{ij}\dot{u}^i\dot{u}^j$ is ϵ, $\epsilon = \pm 1$, and if the curvature of β at P is not 0, then the torsion of β at P is $\epsilon(-K)^{\frac{1}{2}}$ and the derivative $\dot{\boldsymbol{n}}$ of $\boldsymbol{n}$ along β has at P the direction of $\epsilon\dot{\boldsymbol{x}} \times \boldsymbol{n}$.*

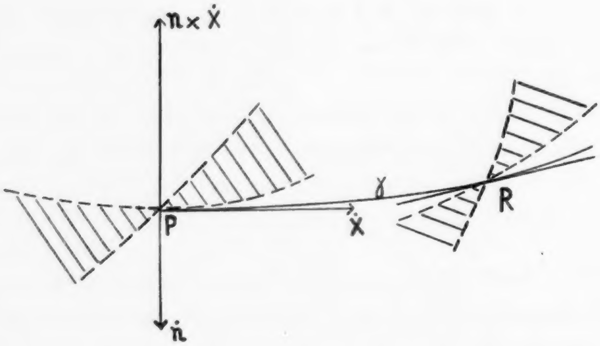
This result will be used in the proof of the following statement.

II. *Let that part of the intersection of S and its tangent plane at P which is in a small angular vicinity of the asymptotic vector $\dot{\boldsymbol{x}} = \boldsymbol{x}_i \dot{u}^i$, be a curve γ of class C^2 and let γ have a non-vanishing curvature in a vicinity of P (of course not at P). Then the asymptotic curve tangent to $\dot{\boldsymbol{x}} = \boldsymbol{x}_i \dot{u}^i$ is separated by the curve γ from its tangent at P.*

Changing, if necessary, the orientation of the space and of the surface, one may suppose that the above-defined signature ϵ is 1, and that the geodesic curvature of γ is > 0.

Let R be a point of γ near P. The position of the tangent plane at R is determined by the position of $\boldsymbol{n}$ at R, and the latter is determined with sufficient accuracy by the fact that $\dot{\boldsymbol{n}}$ has the direction of $\dot{\boldsymbol{x}} \times \boldsymbol{n}$ at P. From the fact that the curvature of γ at R is positive, it follows that γ is on the negative side of the tangent plane to S at R. Thus the asymptotic direction through R, which corresponds to $\epsilon = 1$, is obtained from the tangent direction to γ at R by a small positive rotation. Hence, the unique solution through P of the differential equation which determines the asymptotic curves of positive torsion, is on the side of γ which is determined by the vector $\boldsymbol{n} \times \dot{\boldsymbol{x}}$. On the other hand, since the curvature of γ is positive (except at P), the curve γ is on that side of its tangent $\dot{\boldsymbol{x}} = \boldsymbol{x}_i \dot{u}^i$ at P, which is determined by the direction of $\boldsymbol{n} \times \dot{\boldsymbol{x}}$. This completes the proof of II.

In the figure, the broken boundary of the shaded regions denotes the asymptotic curves, while the shaded regions themselves correspond to the signature $+1$ of the form $h_{ij}\dot{u}^i\dot{u}^j$.



Note that, as a consequence of II, the asymptotic curve through P and tangent to γ has a flex-point at P, whenever γ combined with its continuation on the other side of P, has a flex-point at P. Thus, even though the curvature of γ at P is 0 (if it exists), it cannot be considered as the general case that γ has a flex point at P.

THE JOHNS HOPKINS UNIVERSITY.

THE DERIVATIVE CIRCULAR CONGRUENCE-REPRESENTATION OF A POLYGENIC FUNCTION.*[1]

By Edward Kasner and John De Cicco.

I. Introduction. We refer to the definitions and notations concerning polygenic functions in the earlier papers. (See the bibliography at the end of this paper.) The main problem of this paper is to study the derivative circular congruence-representation of a polygenic function.

The derivative $\gamma = dw/dz$ of a polygenic function $w = \phi(x, y) + i\psi(x, y)$ with respect to the complex variable $z = x + iy$ establishes an element-to-point transformation T between the lineal elements (x, y, θ) of the z-plane and the points (α, β) of the γ-plane, where $\gamma = \alpha + i\beta$. It has been proved that T is characterized by the following three properties: I. The *circle* property; II. The *ratio* $-2:1$ property; and III. The *affine-similitude* property. This element-to-point transformation T induces a correspondence between the ∞^2 points of the z-plane and a certain set of ∞^2 clocks or circles of the γ-plane. We shall term such a correspondance the *derivative clock or circular congruence-representation of a polygenic function.* Any arbitrary correspondance between the ∞^2 points of the z-plane and a certain set of ∞^2 clocks or circles of the γ-plane is called a *clock or circular congruence-representation.* Such a congruence-representation is not necessarily a derivative congruence-representation of a polygenic function. The problem of this paper is to obtain the analytic conditions necessary and sufficient that a given congruence-representation of clocks or circles be a derivative clock or circular congruence-representation of a polygenic function. We derive the following two conclusions:

(A). *A congruence-representation of clocks pictures the derivative of either none or ∞^2 polygenic functions.*

(B). *A congruence-representation of circles pictures the derivative of either none, ∞^2, $2\infty^2$, or ∞^3 polygenic functions.*

II. The derivative clock congruence-representation of a polygenic function. Before proceeding with our work, it will be convenient to introduce the following two linear operators. Let $\mathfrak{D}$ and $\mathfrak{P}$ denote

$$(1)\qquad \mathfrak{D} = \tfrac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right), \qquad \mathfrak{P} = \tfrac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right).$$

* Received December 27, 1938; Revised May 11, 1939.

[1] Presented to the American Mathematical Society, September, 1938.

These two operators are very important in the Theory of Polygenic Functions. Obviously $\mathfrak{D}$ and $\mathfrak{P}$ are commutative. We call $\mathfrak{D}(w)$ the *mean or center derivative,* and $\mathfrak{P}(w)$ the *phase derivative* of the polygenic function w.

Let a polygenic function $w = \phi(x, y) + i\psi(x, y)$ be an *analytic polygenic function,* that is, let ϕ and ψ be each expansible in a power series. Then w may be expressed as a complex function (power series) of the two complex variables $z = x + iy$ and $\bar{z} = x - iy$. That is, w is of the form $F(z, \bar{z})$. *The mean derivative $\mathfrak{D}(w)$ is equivalent to the formal partial derivative of w with respect to $z = x + iy$, and the phase derivative $\mathfrak{P}(w)$ is equivalent to the formal partial derivative of w with respect to $\bar{z} = x - iy$.* We note that the phase derivative $\mathfrak{P}(w)$ is identical in value to Pompieu's areolar derivative.

It may be shown by means of (1) that the derivative $\gamma = dw/dz$ can be written in the form

$$\gamma = \frac{dw}{dz} = \mathfrak{D}(w) + e^{-2i\theta}\mathfrak{P}(w), \tag{2}$$

where

$$\begin{aligned} H + iK &= \mathfrak{D}(w) = \tfrac{1}{2}(\phi_x + \psi_y) + \frac{i}{2}(-\phi_y + \psi_x), \\ h + ik &= \mathfrak{P}(w) = \tfrac{1}{2}(\phi_x - \psi_y) + \frac{i}{2}(\phi_y + \psi_x). \end{aligned} \tag{3}$$

Since $\mathfrak{D}$ and $\mathfrak{P}$ are commutative, we find by the preceding equations

$$\mathfrak{P}(H + ik) = \mathfrak{D}(h + ik). \tag{4}$$

This is the Property III of the preceding papers.

By means of equations (2) and (3), it is found that *the derivative $\gamma = dw/dz$ is pictured by the congruence-representation of clocks*

$$\begin{aligned} H &= \tfrac{1}{2}(\phi_x + \psi_y), \qquad K = \tfrac{1}{2}(-\phi_y + \psi_x), \\ h &= \tfrac{1}{2}(\phi_x - \psi_y), \qquad k = \tfrac{1}{2}(\phi_y + \psi_x). \end{aligned} \tag{5}$$

Thus to every point (x, y) of the z-plane corresponds by means of $\gamma = dw/dz$ that clock of (5) determined by the particular pair of values (x, y).

In this section, we wish to study the converse problem. Let $H(x, y)$, $K(x, y)$, $h(x, y)$, $k(x, y)$ be four arbitrary real functions of (x, y), and for every point (x, y) in the z-plane, construct in the γ-plane the clock whose central vector is $H + iK$ and whose principal radius vector is $h + ik$. What are the necessary and sufficient conditions to be fulfilled by H, K, h, k, in order that the congruence-representation of clocks so obtained may depict the derivative clock congruence-representation of a polygenic function? In other words, when can we find two functions $\phi(x, y)$ and $\psi(x, y)$ such that the relations (5)

expressing the correspondence between the clocks of the congruence and the points of the z-plane are fulfilled?

In the solution of our problem, we shall suppose that the four functions H, K, h, k possess continuous partial derivatives of the first order with respect to x and y. Now if the congruence-representation of clocks is to depict the derivative clock congruence-representation of a polygenic function $w = \phi + i\psi$, we find from (5) that we must have

$$\phi_x = H + h, \qquad \psi_x = K + k,$$
$$(6) \qquad \phi_y = -K + k, \qquad \psi_y = H - h.$$

In order that ϕ and ψ exist to satisfy the preceding equations, we must have

$$(7) \qquad H_x - K_y = h_x + k_y, \qquad H_y + K_x = -h_y + k_x.$$

These conditions are those of compatibility and they are both necessary and sufficient for the existence of the functions ϕ and ψ. The above two real equations are obviously equivalent to the single complex equation (4). Hence

THEOREM 1. *In order that the congruence-representation of clocks $H(x, y) + iK(x, y)$, $h(x, y) + ik(x, y)$ picture the derivative clock congruence-representation of a polygenic function, it is necessary and sufficient that the four functions H, K, h, k satisfy identically the single complex equation* (4) *(Property III). Thus the four functions H, K, h, k must satisfy two real partial differential equations of the first order in x and y.*

Let now the complex equation (4) or the two real equations (7) be identically satisfied. Then by (6), the functions ϕ and ψ are each determined uniquely except for an additive real constant. Hence $w = \phi + i\psi$ is determined uniquely except for an additive complex constant. Thus we obtain

COROLLARY. *Any two polygenic functions with the same derivative clock congruence-representation differ merely by a complex constant. Thus there are ∞^2 polygenic functions with the same derivative clock congruence-representation.*

III. The derivative circular congruence-representation of a polygenic function. Upon taking the circles of the clocks of the preceding section, we see that *the derivative $\gamma = dw/dz$ is pictured in the γ-plane by the congruence-representation of circles*

$$(8) \qquad (\alpha - H)^2 + (\beta - K)^2 = R^2,$$

where

$$H = \tfrac{1}{2}(\phi_x + \psi_y), \qquad K = \tfrac{1}{2}(-\phi_y + \psi_x),$$
$$R^2 = h^2 + k^2 = \tfrac{1}{4}[(\phi_x - \psi_y)^2 + (\phi_y + \psi_x)^2]. \tag{9}$$

Thus to every point (x, y) of the z-plane corresponds by means of $\gamma = dw/dz$ that circle of (8) determined by the particular pair of values (x, y).

In this section, we wish to study the converse problem. Let $H(x, y)$, $K(x, y)$, $R(x, y)$ be three arbitrary real functions of (x, y) and for every point (x, y) in the z-plane construct in the γ-plane the circle whose center is (H, K) and whose radius is R. What are the necessary and sufficient conditions to be fulfilled by H, K, and R, in order that the congruence-representation of circles so obtained may map the derivative of a polygenic function? In other words, when can we find two functions $\phi(x, y)$ and $\psi(x, y)$ such that the relations (9) expressing the correspondence between the circles of the congruence and the points of the z-plane are fulfilled?

In the solution of the general case, we shall suppose that the three functions H, K, R possess continuous partial derivatives of the third order with respect to x and y. In order that the circles $H(x, y)$, $K(x, y)$, $R(x, y)$ depict the derivative circular congruence-representation of a polygenic function, they must be the base circles of a derivative congruence-representation of clocks. These clocks must be given by the four functions $H(x, y)$, $K(x, y)$, $h = R(x, y) \cos \lambda(x, y)$, $k = R(x, y) \sin \lambda(x, y)$ where $\lambda(x, y)$ is to be determined. For these clocks to be a derivative congruence-representation, it is necessary and sufficient by Theorem 1 that

$$\mathfrak{P}(H + iK) = \mathfrak{D}(h + ik) = \mathfrak{D}(Re^{i\lambda}). \tag{10}$$

Upon setting the real and imaginary parts of this equation equal to zero, we find that this equation is equivalent to the two real equations

$$H_x - K_y = (R_x + R\lambda_y) \cos \lambda - (-R_y + R\lambda_x) \sin \lambda,$$
$$H_y + K_x = (R_x + R\lambda_y) \sin \lambda + (-R_y + R\lambda_x) \cos \lambda. \tag{11}$$

Upon assuming that the circles are proper circles (that is, $R \neq 0$), we can solve these equations for λ_x and λ_y obtaining

$$\lambda_x = \frac{(H_y + K_x)}{R} \cos \lambda - \frac{(H_x - K_y)}{R} \sin \lambda + \frac{R_y}{R},$$
$$\lambda_y = \frac{(H_y + K_x)}{R} \sin \lambda + \frac{(H_x - K_y)}{R} \cos \lambda - \frac{R_x}{R}. \tag{12}$$

These equations are of course equivalent to the equations (11). Now let

$$A + iB = \frac{2}{R^2} \mathfrak{P}(H + iK), \qquad \rho = \log R. \tag{13}$$

Then the equations (12) may be written in the form

$$(14)\qquad \begin{aligned}\lambda_x &= RB\cos\lambda - RA\sin\lambda + \rho_y,\\ \lambda_y &= RB\sin\lambda + RA\cos\lambda - \rho_x.\end{aligned}$$

In order that λ exist to satisfy these equations, it is necessary and sufficient that λ satisfy identically the equation

$$(15)\qquad (A_x - B_y)\cos\lambda + (A_y + B_x)\sin\lambda = -R(A^2+B^2) + \frac{1}{R}(\rho_{xx} + \rho_{yy}),$$

which is obtained by setting the partial derivative with respect to y of the right-hand side of the first of equations (14) equal to the partial derivative with respect to x of the right-hand side of the second of equations (14), and also the equations (14).

Now let

$$(16)\qquad \begin{aligned}C + iD &= 2\mathfrak{P}(A + iB),\\ S &= -R(A^2+B^2) + \frac{1}{R}(\rho_{xx} + \rho_{yy}).\end{aligned}$$

Making these substitutions into (15), we find that (15) can be written in the form

$$(17)\qquad C\cos\lambda + D\sin\lambda = S.$$

Upon assuming that $C + iD \neq 0$, we obtain immediately

$$(18)\qquad e^{i\lambda} = \frac{C + iD}{S + iT},$$

where the vector $S + iT$ is such that its absolute value is equal to that of $C + iD$. (This defines the function T.)

The function λ as determined by equation (18) must satisfy identically the equations (14) which are equivalent to the equation (4). This proves that there are one or two derivative congruence-representations of clocks possible; and also that each of these congruence-representations must possess the Property III.

Of course the analytic conditions are that λ as determined by (18) must satisfy identically the equations (14). Thus the conditions are found to be two real partial differential equations of the third order in x and y, which must be satisfied identically by the three functions $H(x,y)$, $K(x,y)$, and $R(x,y)$. Thus we obtain

THEOREM 2. *In order that the congruence-representation of proper circles* $H(x,y)$, $K(x,y)$, $R(x,y) \neq 0$, *with* $C + iD \neq 0$, *picture the deriva-*

tive circular congruence-representation of a polygenic function, it is necessary and sufficient that the clocks $H + iK$, $h + ik = R(C + iD)/(S + iT)$, *(where* C, D, S *and* T *are given by equations* (13), (16), *and* (18)), *possess the Property III. This means that the functions* H, K, *and* R *satisfy identically two real partial differential equations of the third order in* x *and* y.

Since there are either one or two derivative congruence-representations of clocks which possess these circles as base circles, we obtain by the Corollary to Theorem 1 the following

COROLLARY. *There are one or two derivative congruence-representations of clocks which possess these circles as base circles. Therefore there are* ∞^2 *or* $2\ \infty^2$ *polygenic functions which possess the given derivative congruence-representation of circles.*

As an application of Theorem 2, we can determine among all congruence-representations of circles $H(x,y)$, $K(x,y)$, $R(x,y)$, those congruence-representations of circles of *constant radius* R which belong to two distinct derivative congruence-representations of clocks. Any congruence-representation of circles with the required property must be of the form [2]

$$H + iK = f(z) - R \int \frac{\partial}{\partial z} \exp\,[i\chi_y \pm i \arcsin e^{\chi_x}]\, d\bar{z}, \tag{19}$$

where f is a monogenic function of $z = x + iy$, χ is a real harmonic function (not an affine linear function), and R is the constant radius of the congruence-representation of circles. Moreover the polygenic functions which possess this derivative congruence-representation of circles are given by

$$w = \int f(z)\, dz - R \int \exp\,[i\chi_y \pm i \arcsin e^{\chi_x}]\, d\bar{z} + C, \tag{20}$$

where C is an arbitrary complex constant. Thus there are $2\ \infty^2$ polygenic functions w which possess the given derivative congruence-representation of circles (19).

We shall not give the proof of this result. A complete discussion will be found in the book, *The Geometry of Polygenic Functions,* which is being written by the authors.

IV. The degenerate cases of section III. In this section, we shall consider the degenerate cases of the preceding section. First let us consider the case where the circles are proper but where $C + iD = 0$. Then by equa-

[2] By the function exp u, we mean the exponential function e^u.

tion (17), we see that in order that the equations (14) be compatible, it is necessary and sufficient that $S = 0$. Hence

THEOREM 3. *In order that the congruence-representation of proper circles $H(x,y)$, $K(x,y)$, $R(x,y) \neq 0$, which satisfy identically the complex partial differential equation of the second order with respect to x and y,*

$$\frac{1}{4}(C + iD) = \mathfrak{P}\left\{ \frac{1}{R^2}\mathfrak{P}(H + iK) \right\} = 0, \tag{21}$$

picture the derivative circular congruence-representation of a polygenic function, it is necessary and sufficient that H, K, and R satisfy identically the equation

$$\tfrac{1}{4}R^3S = -[\mathfrak{P}(H + iK)][\mathfrak{D}(H - iK)] + R^2\mathfrak{P}\mathfrak{D}\log R = 0. \tag{22}$$

This equation shows that H, K, and R must satisfy a single real partial differential equation of the second order in x and y.

Let now the equations (21) and (22) be identically satisfied by the functions H, K, and $R \neq 0$. Then the condition of compatibility (15) or (17) for the function λ is identically satisfied. By (14), it follows that the complete solution for λ contains an arbitrary real constant. This means that there are ∞^1 derivative congruence-representations of clocks which possess the given congruence-representation of circles as base circles. Then by the Corollary to Theorem 1, there are ∞^3 polygenic functions which possess this given derivative congruence-representation of circles. Thus we have proved the following

COROLLARY. *The circular congruence-representation of Theorem 3 are the base circles of ∞^1 derivative congruence-representations of clocks. From this, it follows that there are ∞^3 polygenic functions which possess this same derivative congruence-representation of circles.*

We shall now give an example of a congruence of circles which actually satisfies the conditions of Theorem 3. Let the circles of Theorem 3 have constant radius. Then the conditions (21) and (22) are together equivalent to the condition

$$\mathfrak{P}(H + iK) = 0, \tag{23}$$

that is, the center transformation must be direct conformal. Then it is found that the polygenic functions which possess these circles as derivative circles are

$$(24) \qquad w = \int (H + iK)\,dz + Re^{i\lambda}(x - iy) + C,$$

where $H + iK$ is of course a monogenic function of $z = x + iy$, R is the constant radius of the congruence of circles, λ is an arbitrary real constant, and C is an arbitrary complex constant. We notice that this is a particular type of harmonic polygenic function.

For since the mean derivative $H + iK$ is a monogenic function of $z = x + iy$, then the phase derivative $h + ik$ by (4) must be a monogenic function of $\bar{z} = x - iy$. But $h^2 + k^2 = R^2$ is constant so that the monogenic function $h + ik$ of $\bar{z} = x - iy$ has constant absolute value. Hence $h + ik = Re^{i\lambda}$ where λ is an arbitrary real constant. From the fact that $H + iK$ is a monogenic function of $z = x + iy$ and the fact that $h + ik$ is the constant $Re^{i\lambda}$, we obtain immediately (24).

Finally let us consider the case where the circles are all null. That is, $R = 0$. In order that the null circles $H(x, y)$, $K(x, y)$, $R = 0$, depict the derivative circle congruence-representation of a polygenic function, they must be the base circles of a derivative congruence-representation of clocks. These clocks must be given by the four functions $H + iK$, $h + ik$, where h and k satisfy the condition $h^2 + k^2 = R^2 = 0$. Hence $h = k = 0$. Thence applying Theorem 1 to these clocks, we find that the center function $H(x, y) + iK(x, y)$ must be direct conformal. Thus

THEOREM 4. *In order that the congruence-representation of null circles $H(x, y)$, $K(x, y)$, $R = 0$, pictures the derivative circular congruence-representation of a polygenic function, it is necessary and sufficient that the center function $H(x, y) + iK(x, y)$ be a monogenic function of $z = x + iy$. Thus the functions H and K satisfy two real partial differential equations of first order in x and y (of course these are the Cauchy-Riemann equations).*

From this theorem and from what was said above, we easily derive

COROLLARY. *The polygenic functions which possess this derivative null circular congruence-representation are monogenic functions of $z = x + iy$, and any two of them differ merely by a complex constant. Thus there are ∞^2 polygenic functions which possess the same derivative null circular congruence-representation.*

Our object in stating this simple theorem which deals essentially with the integral of a monogenic function is to contrast this with the complicated situations which arise in the general case where the circles are not null circles.

The results stated in Theorems 2 and 3, therefore, may be regarded as generalizations of a classic theorem of integration.

COLUMBIA UNIVERSITY,
BROOKLYN COLLEGE.

BIBLIOGRAPHY.

We refer to the earlier papers by Kasner:

1. "A new theory of polygenic (or non-monogenic) functions," *Science*, vol. 66 (1927), pp. 581-582.

2. "General theory of polygenic or non-monogenic functions. The derivative congruence of circles," *Proceedings of the National Academy of Sciences*, vol. 14 (1928), pp. 75-82.

3. "Note on the derivative circular congruence of a polygenic function," *Bulletin of the American Mathematical Society*, (1928), pp. 561-565.

4. "A complete characterization of the derivative of a polygenic function," *Proceedings of the National Academy of Sciences*, vol. 22 (1936), pp. 172-177.

5. "Polygenic functions whose associated element-to-point transformation converts unions into points," *Bulletin of the American Mathmatical Society*, (1938), pp. 726-732.

6. "Géométrie des fonctions polygènes," *Atti del Congresso Internazionale dei Matematici*, Bologna (1928), pp. 255-260.

Also see abstract 44-7-289, "Additional properties of the second derivative of a polygenic function," and abstract 45-1-84, "the polygenic functions whose associated series are equiparallel series" by De Cicco, *Bulletin of the American Mathematical Society*.

THE CANONICAL LINES AND THE EXTREMALS OF TWO INVARIANT INTEGRALS.*

By RUTH B. RASMUSEN.

1. Introduction. In the first part of this note it is shown by means of the asymptotic osculating quadrics defined for the union curves of a canonical line congruence and certain quadrics of Darboux that the canonical lines of the first kind may be associated in pairs as cusp-axes of cones of the third class. Dually, the canonical lines of the second kind may be associated in pairs as flex-rays of cubics which lie in the tangent plane of the surface. In the second part the extremals of two well known invariant integrals are interpreted geometrically.

2. Analytic basis. This section summarizes portions [1] of the classical analytical theory of the projective differential geometry of surfaces which are used in later developments. In ordinary space in which a point has projective homogeneous coördinates $x^{(1)}, \cdots, x^{(4)}$, the parametric vector equation of an analytic non-ruled surface is

$$x = x(u, v), \tag{1}$$

the parameters being u, v. If the asymptotic curves on the surface are the parametric curves, the coördinates x satisfy a system of two partial differential equations which can be reduced to the form

$$\begin{aligned} x_{uu} &= px + \theta_u x_u + \beta x_v, \\ x_{vv} &= qx + \gamma x_u + \theta_v x_v \qquad (\theta = \log \beta\gamma), \end{aligned} \tag{2}$$

subscripts indicating partial differentiation and the coefficients being functions of u, v which satisfy certain integrability conditions.

If a curve C on the surface is regarded as imbedded in the one-parameter family of curves defined by the equation

$$dv - \lambda du = 0, \tag{3}$$

where λ is a function of u, v, the local equation of the osculating plane at a point P_x of the curve C defined by equation (3) is

* Received August 20, 1938; Revised December 9, 1938.

[1] This material may be found in Chapter III and in the exercises at the end of Chapter III of Lane's *Projective Differential Geometry of Curves and Surfaces*, The University of Chicago Press, 1932.

$$2\lambda(\lambda x_2 - x_3) + (\lambda' + \beta - \theta_u\lambda + \theta_v\lambda^2 - \gamma\lambda^3)x_4 = 0, \tag{4}$$

where $\lambda' = \lambda_u + \lambda\lambda_v$.

The equations of the asymptotic osculating quadrics Q_u and Q_v of the curve C at a point P_x of a surface are

$$\begin{aligned} 2\lambda^3(x_2x_3 - x_1x_4) + 2\beta\lambda x_4(x_3 - \lambda x_2) + \{\beta[\lambda' - \beta \\ + (\phi - \theta_u)\lambda - (2\psi - \theta_v)\lambda^2] - (\beta\gamma + \theta_{uv})\lambda^3\}x_4^2 = 0, \end{aligned} \tag{5}$$

$$\begin{aligned} 2(x_2x_3 - x_1x_4) - 2\gamma\lambda x_4(x_3 - \lambda x_2) + \{\gamma[-\lambda' - \gamma\lambda^3 \\ + (\psi - \theta_v)\lambda^2 - (2\phi - \theta_u)\lambda] - (\beta\gamma + \theta_{uv})\}x_4^2 = 0. \end{aligned} \tag{6}$$

The osculating plane (4) at a point P_x of a curve C of the family (3) on a surface intersects the quadric (5) in a conic. The locus of this conic as the curve C varies but remains tangent to a fixed line l at P_x is the quadric

$$\begin{aligned} 2\lambda^3(x_2x_3 - x_1x_4) + 4\beta\lambda x_4(x_3 - \lambda x_2) + [\beta(-2\beta + \phi\lambda \\ - 2\psi\lambda^2) - \theta_{uv}\lambda^3]x_4^2 = 0, \end{aligned} \tag{7}$$

where

$$\phi = (\log \beta\gamma^2)_u, \qquad \psi = (\log \beta^2\gamma)_v. \tag{8}$$

When the quadric (6) is used in place of the quadric (5), the corresponding locus is the quadric

$$\begin{aligned} 2(x_2x_3 - x_1x_4) - 4\gamma\lambda x_4(x_3 - \lambda x_2) + [\gamma(-2\gamma\lambda^3 \\ + \psi\lambda^2 - 2\phi\lambda) - \theta_{uv}]x_4^2 = 0. \end{aligned} \tag{9}$$

At a point P_x of a surface the quadric of Darboux,

$$2(x_2x_3 - x_1x_4) - [\beta\gamma(1 - k) + \theta_{uv}]x_4^2 = 0, \quad (k = \text{a constant}) \tag{10}$$

is the quadric of Lie if $k = 0$, the canonical quadric of Wilczynski if $k = 1$. If any line l meets this quadric at P_x and also at P_k, if the line l meets the quadric of Lie at P_0 and meets the quadric of Wilczynski at P_1, then

$$(P_xP_0P_1P_k) = k,$$

so that by means of this cross ratio the quadric of Darboux is characterized geometrically for an arbitrary value of k.

The equation of the canonical plane of the surface at the point P_x is

$$\phi x_2 - \psi x_3 = 0. \tag{11}$$

The pencil of lines lying in this plane and having its center at P_x is called the first canonical pencil of the surface at the point P_x, and any line of this pencil joining P_x to a point

$$(12) \qquad (0,\ k\psi,\ k\phi,\ 1) \qquad (k = \text{a constant})$$

is spoken of as a canonical line of the first kind.

The curves defined on a surface by a differential equation of the form

$$(13) \qquad v'' = A + Bv' + Cv'^2 + Dv'^3,$$

in which the coefficients are functions of u, v and accents indicate total differentiation with respect to u, are called hypergeodesics.

3. The canonical lines of the first kind as cusp-axes. The differential equation of the union curves of a canonical line congruence is an equation of the form (13) with

$$(14) \qquad A = -\beta, \quad B = \theta_u + 2r\phi, \quad C = -(\theta_v + 2r\psi), \quad D = \gamma,$$

where k in (12) has been replaced by r.

The equation of the asymptotic osculating quadric Q_u at a point P_x defined for a union curve C of a canonical line congruence is

$$(15) \qquad \begin{aligned} &2\lambda^3(x_2x_3 - x_1x_4) + 2\beta\lambda x_4(x_3 - \lambda x_2) \\ &\quad + \{\beta[-2\beta + \lambda\phi(2r+1) - 2\lambda^2\psi(r+1)] - \theta_{uv}\lambda^3\}x_4^2 = 0. \end{aligned}$$

The residual conic of intersection of the quadric (15) and the quadric of Lie lies in the plane

$$(16) \quad 2\lambda(x_3 - \lambda x_2) + \{-2\beta + 2\lambda\phi(2r+1) - 2\lambda^2\psi(r+1) + \gamma\lambda^3\}x_4 = 0.$$

When the quadric (6) is used in place of (5), the equation of the corresponding plane is

$$(17) \quad 2\lambda(x_3 - \lambda x_2) - [\beta - 2\lambda\phi(r+1) + \lambda^2\psi(2r+1) - 2\gamma\lambda^3]x_4 = 0.$$

The harmonic conjugate of the tangent plane of the surface at the point P_x with respect to the planes (16) and (17) is

$$(18) \quad 4\lambda(x_3 - \lambda x_2) + [-3\beta + 3\gamma\lambda^3 + \lambda\phi(4r+3) - \lambda^2\psi(4r+3)]x_4 = 0.$$

To find the equation in plane coördinates of the cone enveloped by the plane (18) we eliminate ρ and λ homogeneously from

$$\begin{aligned} &\rho u_1 = 0, \qquad \rho u_2 = -4\lambda^2, \qquad \rho u_3 = 4\lambda, \\ &\rho u_4 = -3\beta + 3\gamma\lambda^3 + \lambda\phi(4r+3) - \lambda^2\psi(4r+3) \end{aligned}$$

to obtain

$$u_1 = u_2u_3[4u_4 - (4r+3)\phi u_3 - (4r+3)\psi u_2] - 3\beta u_3^3 - 3\gamma u_2^3 = 0.$$

This cone has for cusp-axis the line joining the point P_x to

$$(19) \qquad [0,\ -(4r+3)\psi/4,\ -(4r+3)\phi/4,\ 1.]$$

Consequently, we have shown that *there is associated with a given canonical line a canonical line which is the cusp-axis of a cone of class three.*

This relation between two canonical lines is symmetric. For, if $\bar{r} = -(4r+3)/4$, then $r = -(4\bar{r}+3)/4$, where $\bar{r}$ is also a constant.

For example, the canonical lines for which k has the following values play symmetric rôles as indicated above: 0 and $-3/4$; $-1/4$ and $-1/2$; $-1/3$ and $-5/12$; etc.

4. The canonical lines of the second kind as flex-rays. Comparing equation (18) with equation (4), one easily verifies that equation (18) is the equation of the osculating planes of all the hypergeodesics (13) through the point P_x for which

$$A = \beta/2, \quad B = \theta_u - (4r+3)\phi/2, \quad C = -\theta_v + (4r+3)\psi/2, \quad D = -\gamma/2.$$

Lane has shown that the locus [2] of the ray-points, corresponding to the point P_x, of all the hypergeodesics (13) that pass through P_x is a cubic curve whose line of inflexions is the reciprocal of the cusp-axis of the cone enveloped by the osculating planes of the hypergeodesics (13). Applying his results to our case, the equation of the cubic becomes

$$x_4 = x_2x_3[2x_1 + (4r+3)\phi x_2/2 + (4r+3)\psi x_3/2] - \beta x_2^3/2 - \gamma x_3^3/2 = 0.$$

The flex-ray of this cubic is a canonical line of the second kind.

5. The extremals of two invariant integrals. The extremals [3] of the invariant integral $\int(\beta/v')\,du$ are hypergeodesics for which

$$A = D = 0, \quad B = (\log\beta)_u/2, \quad C = (\log\beta)_v.$$

Let us substitute these values of A, B, C, D in equation (13), replacing v' by λ and v'' by λ'. Inserting the value of λ' thus obtained in equation (4), we obtain the equation

$$(20) \qquad 2\lambda(\lambda x_2 - x_3) + [\beta - \gamma\lambda^3 + \lambda^2\psi - \lambda\phi/2]x_4 = 0.$$

The equation of the plane which contains the residual conic of intersection of the quadric (7) and the quadric (10) with $k = -1$ is identical with the plane (20). So we have proved the theorem:

A curve on a surface is an extremal of the invariant integral $\int(\beta/v')\,du$ if, and only if, at each of its points its osculating plane contains the residual conic of intersection of the quadric (7) *and the quadric* (10) *with* $k = -1$.

---

[2] Lane, *loc. cit.*, p. 101.
[3] Lane, *loc. cit.*, p. 117.

The residual conic of intersection of the quadric (7) and the quadric of Lie lies in the plane whose equation is

$$(21)\qquad 4\lambda(x_3 - \lambda x_2) + [-2\beta + \phi\lambda - 2\psi\lambda^2 + \gamma\lambda^3]x_4 = 0.$$

The residual conic of intersection of the quadric (7) and the quadric of Wilczynski lies in the plane whose equation is

$$(22)\qquad 4\lambda(x_3 - \lambda x_2) + [-2\beta + \phi\lambda - 2\psi\lambda^2]x_4 = 0.$$

An easy calculation shows that the harmonic conjugate of the plane (22) with respect to the tangent plane of the surface at the point P_x and the plane (21) is the plane (20). Therefore, *a curve on a surface is an extremal of the invariant integral* $\int(\beta/v')du$ *if, and only if, at each of its points its osculating plane is the harmonic conjugate of the plane* (22) *with respect to the tangent plane of the surface at the point* P_x *and the plane* (21).

The reader will observe that the above characterizations are essentially equivalent.

The extremals of the invariant integral $\int \gamma v'^2 du$ which are hypergeodesics for which

$$A = D = 0, \quad B = -(\log\gamma)_u, \quad C = -(\log\gamma)_v/2,$$

may be characterized by a method similar to the one employed above. The equation of the osculating plane of these extremals is

$$(23)\qquad 4\lambda(x_3 - \lambda x_2) + [-2\beta + 2\phi\lambda - \psi\lambda^2 + 2\gamma\lambda^3]x_4 = 0.$$

The harmonic conjugate of the tangent plane to the surface at the point P_x with respect to the planes (20) and (23) is

$$8\lambda(x_3 - \lambda x_2) + [-4\beta + 3\phi\lambda - 3\psi\lambda^2 + 4\gamma\lambda^3]x_4 = 0.$$

It is easily verified that *this plane envelops a cone of class three whose cusp-axis is the canonical line for which* $k = -3/8$. This interpretation of the canonical line[4] for which $k = -3/8$ is believed to be new.

WILSON JUNIOR COLLEGE,
CHICAGO, ILLINOIS.

[4] For another definition see Lane, *loc. cit.*, p. 117.